DOCUMENT RESUME

ED 092 640 UD 014 248

TITLE	An Evaluation of the 1972 Bilingual Education Program, ESEA Title VII, Community School District 9, Bronx, New York.
INSTITUTION	Mobicentrics Inc., Bronx, N.Y.
SPONS AGENCY	Community School District 9, Bronx, N.Y.
PUB DATE	Jul 73
NOTE	182p.; Function 49-37601
EDRS PRICE	MF-$0.75 HC-$9.00 PLUS POSTAGE
DESCRIPTORS	*Bilingual Education; Bilingual Students; Curriculum Development; Disadvantaged Youth; *Elementary Schools; Inservice Teacher Education; Instructional Materials; Paraprofessional School Personnel; Preservice Education; *Program Evaluation; Spanish Speaking; Urban Schools
IDENTIFIERS	Elementary Secondary Education Act Title VII; ESEA Title VII; *New York City

ABSTRACT

This program, funded under Title VII of the Elementary Secondary Education Act of 1965, was designed to provide an open learning environment in which students could learn to function in both English and Spanish. The classrooms were large and for the most part contained several grades which were divided into separate classes by means of dividers. The program operated with a staff of 15 teachers, 1 director, and 6 paraprofessionals for 365 students. The activities of the program included instruction in both dominant and secondary languages, culture and heritage, math, social studies, reading and language arts. The primary objectives of the program were: (1) the mean scores of students on various tests should significantly increase between pre- and posttesting; (2) teachers would attend preservice and inservice training in the teaching of language arts, math, and reading; (3) teachers would enroll in a bilingual education program at a metropolitan university; (4) academic materials would be developed and acquired; and (5) an exchange of materials with a sister school in Puerto Rico would be set up. The primary medium of instruction was the students' dominant language, English. There appeared to be a marked increase in Spanish as a medium of instruction over the year. (Author/JM)

ED 092640

49-37601

An Evaluation of the 1972 Bilingual Education Program ESEA Title VII Community School District 9 Bronx, New York

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

For:
District 9
Community School Board of
The New York City Board of Education

An evaluation of a New York City school district educational project funded under Title VII of the Elementary and Secondary Education Act of 1965 (PL 89-10) performed under contract with the Board of Education of the City of New York for the 1972-73 school year.

UD 014240

mobicentrics inc, 384 east 149th street, bronx, new york 10455 • 993-1000

JULY 1973

ED 092610

49-37601

An Evaluation of the 1972 Bilingual Education Program ESEA Title VII Community School District 9 Bronx, New York

For:
District 9
Community School Board of
The New York City Board of Education

An evaluation of a New York City school district educational project funded under Title VII of the Elementary and Secondary Education Act of 1965 (PL 89-10) performed under contract with the Board of Education of the City of New York for the 1972-73 school year.

inobicentrics inc., 384 east 149th street, bronx, new york 10455 • 993-1000

JULY 1973

TABLE OF CONTENTS

		PAGE
I.	EXECUTIVE SUMMARY	1
II.	PROGRAM DESCRIPTION	4
III.	EVALUATION METHODOLOGY	8
IV.	FINDINGS	10
	A. Pupil Achievement	10
	B. Classroom Structure, Content and Language	167
	C. Staff Development	175
	D. Community Activities	177
	E. Materials	178
V.	RECOMMENDATIONS	179

I. EXECUTIVE SUMMARY

A. PROGRAM DESCRIPTION

The bilingual program at the Rafael Hernandez School, C.E.S. 235 operated with a staff of 15 teachers, one director and 6 paraprofessionals for 365 students. The concept of the school was to provide an open learning environment in which students could learn to function in both English and Spanish language. The classrooms were large and for the most part contained several grades which were divided into separate classes by means of dividers. The activities of the program included instruction in both dominant and secondary languages, culture and heritage, math, social studies, reading and language arts. The primary objectives of the program were:

- The mean scores of students on various tests should significantly increase between pre and post-testing.
- The teachers will attend pre-service and in-service training in the teaching of language arts, math and reading.
- Teachers will have enrolled in a bilingual education program at a Metropolitan University.
- Academic materials were to be developed and acquired.
- An exchange of materials with the sister school was to be set up.

B. SUMMARY OF FINDINGS

These findings relate to the specific objectives outlined in the following chapter. More detailed information on data collected can be found in the "Findings" Chapter of this report.

1. Instructional Component

- The primary medium of instruction was the students dominant language, English. There does seem though to be a marked increase in Spanish as a medium of instruction between this year and last year.
- The mean score of children in the second through sixth grades tested in vocabulary was almost perfect in English, but far lower in Spanish. In phonemic discrimination, children scored low in both Spanish and English In English the change from pre to post-test was positive and significant (beyond the .05 level) in all grades but third. In Spanish, however, there was a significant

negative change or deterioration between pre and post-test.

- The mean score for kindergarten and first grades was almost perfect for English and met or came very close to the criteria of achieving the 50th percentile in both pre and post-test for both grades. In Spanish, however, children fell well below the established norms. We can conclude from this that the sample was not not totally bilingual, but was certainly moving in that direction.

- There was a significant improvement in vocabulary on the Inter-American: Test of Reading in second and third grade only. In Speed of Comprehension there was no significant gain in any grade except third. In relation to the national sample, the children scored below grade norms. In Level of Comprehension, children in every grade significantly improved during the school year. In addition in every grade except fourth, they scored above or near the 50th percentile rank for their grade.

- There was a significant improvement in Word Knowledge in every grade. In Reading as well, all grades showed significant improvement, however, only in second, fifth and sixth grades did students achieve beyond what they were expected to on the post-test.

2. Staff Development Component

- Pre-service training was conducted for all new teachers and old teachers received continuing in-service training. Well over 3/4 of the staff attended training sessions and the attendance rate was from 90-100% at meetings.

- All teachers except one (well over 80%) had enrolled in courses at one of the Universities in the New York area and had amassed nine credits of work.

3. Materials Component

- Academic materials were acquired and developed.

- No exchange of materials with the sister school in Puerto Rico has yet been established.

- Materials (books and articles) for a Professional Staff Library were acquired.

C. SUMMARY OF RECOMMENDATIONS

- More emphasis should be placed on English Vocabulary development in every grade except sixth.
- More emphasis needs to be placed on reading skills in all grades.
- A greater effort should be made in making the classroom environment reflect a bilingual culture. Only 42% of the classes observed attempted to do this.
- More time should be spent teaching the Spanish language.
- Greater use of paraprofessionals to increase the flexibility of the classroom organization should be attempted.
- In general, there is a substantial improvement over last year in almost every area. The evaluators recommend that the program definitely be recycled next year, as the program appears to be achieving beneficial effects on the students involved. The school appears to be moving toward a truly bilingual open classroom situation.

PROGRAM DESCRIPTION

A. OBJECTIVES

The evaluation of CES 235 for the year 1972-1973 focused on the three main program components, an instructional component, a staff development component and a materials component, in contrast to the evaluation done in 1971-1972 which looked at other components as well.[1] The specific objectives of these components as mentioned in the program proposal follow.

1. Instructional Component

- The primary medium of instruction/learning in all classes will be the students' dominant language; however, there will be a progressive decrease in the use of the dominant language as the medium of instruction as one moves from K to grade four.

- The mean score of the students in grades 2 - 6 tested in their dominant language will be at or above the 50th percentile norms for their respective ages by the end of the school year on the sub-tests of the Tests of Basic Language Competence, Level II. There will be a significant increase between pre and post-test.

- The mean score for the kindergarten and first grade children tested in their dominant language will be at or above the 50th percentile means for their ages by the end of the school year on Subtest No. 1, Oral Vocabulary of the Tests of Basic Language Competence, Level I. The 50th percentile raw score for kindergarten is 20 and for first grade 22. No performance level is specified for second language. There will be a significant improvement between pre and post-test.

- The mean score of the Spanish-dominant students in grades 1 - 6 and of the English-dominant students will be at or above the 50th percentile norm for their age groups by the end of the school year as measured by the Inter-American Series: Test of Reading or Prueba de Lectura, Levels I, II and III. No performance is specified for second language reading ability. There will be a significant increase between pre and post-test.

---

[1] See Evaluation of Title VII, Bilingual Education Program, Community School District 9, Bronx, N.Y., 1971-1972, by Mobicentrics, Inc.

- There will be a significant improvement between pre and post-test in reading ability as measured by the Metropolitan Achievement Tests. The test will be given to grades 2 to 6 but not to the Spanish dominant pupils who are rated below C on the N.Y.C. Scale of Ability to Speak English.

2. Staff Development Component

- By September 30, 1972, initial training in the Gattegno method will have been conducted for new teachers. By September 30, 1972 the returning staff and the new teachers will have received training in the adaption of the informal education theory to the bilingual setting. It is expected that at least 3/4 of the bilingual school instructional staff will have attended such training with an attendance rate of 85% or above.

- By June 30, 1973, 90% of the instructional staff (i.e., teachers and paraprofessionals) will have participated in at least 75% of the on-going, in-service training conducted by the supervisory staff consultants.

- By June 30, 1973, 80% of the teachers, enrolled in a bilingual education program at a metropolitan university, will have amassed nine credits of work.

3. Materials Component

- Academic materials relating to English and Spanish Language Arts and Culture were to be acquired and developed.

- An exchange of materials with the sister school in Puerto Rico was to be explored and set up.

- Materials for staff development were to be acquired.

B. STUDENTS

Elementary school C.E.S. 235 was established in 1972 to provide a bilingual instructional program in an informal setting to Spanish dominant and English dominant pupils. The pupil population was to total approximately 350, with about 50% of the pupils being Spanish dominant and 50% being English dominant. It was to range from kindergarten through sixth grade. Pupils were to be selected for participation in the C.E.S. 235 program on the basis of local residency, poor English

language competency, and desire to enter the program.

C. STAFF

The school staff included teachers, paraprofessionals, guidance personnel, consultants for staff training, a curriculum specialists, a bilingual teacher in school and community relations, a director and an assistant director. All staff was bilingual.

D. ACTIVITIES

The instructional component focused on language arts and culture and heritage taught through the dominant language. Language arts included English as a second language and Spanish as a dominant language for Spanish dominant pupils and Spanish as a second language and English as a dominant language for English dominant pupils. The content areas included math, social studies, science, art and music. Culture and heritage were taught in conjunction with the regular social studies program.

These skills were nurtured in an informal, nongraded, open classroom setting which responded to the human needs of the child. This open classroom environment was initiated in the first year of this program in the lower grades. The instructional program emphasized an integrated curriculum in which the various subject areas were related to each other. The instructional program emphasized its bilingual nature, by using both languages for instruction and allowing children with different skills in each language to teach and act as language models for each other.

Staff development was one of the principal objectives for the first year of bilingual school C.E.S. 235. It was to be provided by means of curriculum and orientation workshops, training and development sessions, administrative meetings, and university courses. The goals of staff development were increased instructional skill in Spanish and English language arts (both as a first and second language) as well as other content areas and increased skills and understanding of the cultural heritage and self-image of the pupils.

C.E.S. 235 was to attempt during its first year to acquire, adapt, and/or develop academic materials relating to English and Spanish language arts, content areas, and the cultural heritage of the children. In addition, materials were to be obtained for staff development. These materials were to make up a professional library. An exchange of materials with the sister school in Puerto Rico was to be explored during

the first year, and a culture and media center in each of the classrooms was to be set up.

III. EVALUATION METHODOLOGY

This section serves as an overall view of the evaluation design for the program. In addition, the design for each component is explained in greater detail in the "Findings" section of this report.

The main goal of the evaluation is to determine to what extent the program as it is administered by the district and implemented in the schools fulfills the activities and objectives set forth in the program proposal.

A. SAMPLE SELECTION

It was decided when the program started in February of 1972, that all the children in the school at a particular point in time would be evaluated, and, therefore, no sampling design was needed. All the students in the school from grade 1 to 6 were tested during the first week of May, 1972. All the students who were still attending the school, were tested again during the first week of June, 1973. For the purpose of our evaluation the data from the tests given in 1972 will be considered as the pre-test data and the data from 1973 will be the post-test data. The design of the evaluation in this way causes some technical problems. The samples in the two testing situations will not be completely independent samples because the majority of the children tested in 1973 would have been in the school in 1972; but on the other hand, the samples will not be completely dependent either since some of the students tested in 1972 will not be in the school anymore; other students to be tested in 1973 were not in the program in 1972.

The most elegant way of solving this technical problem would be by selecting for the evaluation only those students for whom there was pre and post-test data. This procedure would be technically flawless but would require a tremendous amount of time and effort. Many children should be rejected from the evaluation because they were not in the program in 1972 or because they were in the program in 1972 but were not in 1973. The approach was considered too stringent and, therefore, was not followed.

Another possibility was to consider the pre and post-test samples as independent samples. Using this procedure we violate to some extent the assumptions underlying the statistical test because the samples are assumed to be completely independent while in fact, they are not; but the analysis is not completely illegitimate because the samples are not identical. This procedure will be followed in our evaluation. Comparisons will be made between mean pre-test scores and mean post-test scores and the t-test for independent samples will be used to test the significance of the difference between the means.

B. INSTRUMENTS

1. Tests

The Metropolitan Achievement Tests, Reading, were given to all students from kindergarten to grade 6 in 1972 and were given again at the end of the program in 1973. It was decided to give the same level for the post-test that the children had received for the pre-test. Kindergarten and first grades received the Primer, second grade, Primary I, third grade, Primary II, fourth and fifth grades, Elementary and the sixth grade the Intermediate level. Only those students who could understand English were tested.

The Inter-American Series Test of Reading was given to grades 1 through 6. Level I was used with first and second grade children, Level II with third grade and Level III with the fourth, fifth and sixth grades. The test was given in Spanish or English according to the children's dominant language.

The Language Competence Test was given to children from kindergarten to sixth grade; children were tested both in Spanish and in English. Subtest No. 1, Oral Vocabulary was given to all grades; Level I was used in kindergarten and Level II in all other grades. Subtest No. 2, Phonemic Discrimination was given to third through sixth grades.

Other tests were given in the 1971-72 school year, but this data was reported for that year.

2. Interview

An interview was conducted with the principal of school C.E.S. 235 in order to determine his perceptions of the school and the main problems encountered in addition to providing ongoing information about the program.

C. OBSERVATIONS

Observations in each class were conducted approximately four times during the year to determine the instructional content of the classroom, the organization of the classroom, the role of paraprofessionals and the language used by children and teachers.

D. ANALYSIS OF RECORDS

Records on demographic background of staff, monthly meetings and other variables were examined.

IV. FINDINGS

A. PUPIL ACHIEVEMENT

1. Evaluation Methodology

Children were tested on the following achievement tests: Metropolitan Achievement Tests, Reading; Inter-American Series, Test of Reading; and Basic Language Competence. More information about each test follows in this report.

Children were tested by their teachers in the first week of May, 1972 and again in the first week of June, 1973 but this time by Mobicentrics specially trained consultants.

Two teacher training sessions concerning administration of these tests were held on April 20 and April 27, 1972. At the first session, each teacher was given a list of the tests for his class and the manuals for each test. Each test was discussed briefly to orient the teachers, and Mobicentrics explained the purpose of the testing. A week was scheduled between the two sessions to allow teachers to study the manuals. At the second session, Mobicentrics read through and discussed the directions for all the tests with the teachers. At this session, teachers were also given a testing schedule, developed by Mobicentrics to ensure that each class took the appropriate tests.

a. Metropolitan Achievement Tests, Reading

Metropolitan Achievement Tests are a series of measures designed to tell teachers and administrators how much pupils have learned in the important content and skill areas of Reading. The tests are basically a source of information yielding comparable measures from one level to another and one subject to another. These features help educators in evaluating pupil progress over the years and in identifying particular strengths and weaknesses. More specifically, the tests, including supplementary interpretive materials, are designed (1) to help teachers plan for instruction based on pupils' needs and differences and to evaluate the effects of previous instruction; and (2) to help administrators assess schoolwide progress toward educational goals and to plan for and evaluate curricular changes.

The test contents are based on careful and extensive analyses of contemporary textbooks, syllabuses, and other curricular sources.

b. Inter-American Series, Test of Reading

The Inter-American Tests are designed for use throughout the Western Hemisphere wherever a test in English or in Spanish may be desired. Both English-speaking and Spanish-speaking educators participated in the test construction to find items requiring activities common to the two cultures, but not necessarily equally common. Indeed, administration of the tests may well reveal cultural differences. The language of the tests was chosen to avoid local idioms, and instead to use "standard" language that could be understood generally. An effort was made also to find language of similar difficulty for the two editions. Since the equivalence of scores on tests in different languages cannot be determined with precision, the statistical results of comparison should always be treated as approximate.

The result is a series of tests in English and another series of tests in Spanish, each valid (1) for use by itself or (2) for use in comparing scores obtained from administration of one series with scores obtained from administration of the other.

The comparability of the English and Spanish editions is the unique contribution of the Inter-American Series of tests, an important asset in a variety of situations. Many schools enroll both Spanish-speaking and English-speaking children, and to these the tests offer special assistance. Each child can be tested in his native language, and the scores in the two languages will be comparable. The tests provide a basis also for comparing the abilities of the same person in the two languages - for example, in schools in which either Spanish or English is taught as a second language. Another contribution of the tests is the assistance they offer to educators in different countries who wish to communicate the results of measurement or to engage in joint research. The tests will serve in different geographical areas as common measuring instruments.

The Tests of Reading in the Inter-American Series are designed to measure both Vocabulary and Comprehension. At levels above grade 1 comprehension is measured by two subtests, Speed of Comprehension and Level of Comprehension.

Scores on the Tests of Reading may be used not only as measures of achievement in reading itself, but also as a basis for estimating ability to do school work in other areas in which the ability to read is related to achievement. In the grouping of children and in adjusting instruction to individual differences, a measure of reading ability is often as useful as a measure of general

mental ability. After a child has learned to read, the use of both measures is much better than the use of one alone.

A measure of the ability to read in one's native language is of special importance to the teacher of a second language. The Inter-American tests enable the teacher to make direct comparisions of achievement in reading English and Spanish.

c. Basic Language Competence Test

The Tests of Basic Language Competence in English and Spanish is an experimental battery of tests designed to measure a basic language competence of children who speak English or Spanish, or who are bilingual in these two languages. It is anticipated that "normal" children who live in New York and who speak English or Spanish (or both languages) as native languages will make near-perfect scores on the sub-tests in the battery. The normal child is to be defined as not mentally or physically defective. Such a normal child may or may not be socially advantaged as well. Children who do not have a basic competence in English or Spanish and who are learning one or the other languages as a second language will make varied scores on these tests, depending on the extent of their basic competence in the second language. Tentative results with tests similar to these suggest that these expectations are justified.

The Test of Basic Language Competence in English and Spanish is a group-administered battery. (Even though the tests have been designed for group administration, it is possible, with only minor changes in the Instructions and Directions, for the test administrator to give these tests individually. If a test is administered individually, the test administrator can have the child himself mark the answer sheet or mark it for him, allowing the child to respond verbally or non-verbally.)

The test appears in two similar versions: English and Spanish. Each version also appears in two different but comparable forms: A and B. In testing bilingual children it is necessary to use different forms of a test if it is given in both languages. Each form of the tests contains six sub-tests: Oral Vocabulary (No. 1), Phonemic Discrimination at the Word and Phrase Levels (No. 2), Phonemic Discrimination at the Sentence Level (No. 3), Grammatical Sensitivity (No. 4), Grammatical Discrimination (No. 5), and Assimilation of Meaning (No.6)

These tests are designed to measure children's basic competence in a language via their perception of linguistic phenomena. The concept of "basic competence"

in a language is here defined to encompass the automatic and unconscious control of the segment of language which serves as stimuli on the tests in one or the other language. This concept must, therefore, be distinguished from the concept of "competence" of the generative-transformational theorists since these tests attempt to measure (i.e., to measure indirectly) only a small, infinite segment of various aspects of true competence, which is practically and logically infinite (at least for the mature adult native speaker).

The tests attempt to measure certain aspects of children's knowledge of language from sound to sense. These tests strive toward being "pure" language tests and they must be distinguished in function from intelligence tests and achievement tests. Of course, there can never be such a thing as a pure language test. This is the case because language competence is only one of a multitude of factors involved in a child's linguistic performance.

In the present evaluation only sub-tests No. 1 and No. 2 were used. Sub-test No. 1, Oral Vocabulary, measures knowledge of a limited number of lexical items, i.e., knowledge of twenty-four concrete nouns. Sub-test No. 2 measures perceptual control of certain phonemic contrasts which are given as stimuli at the word, phrase, and sentence levels and which are under the pressure of interference from the second language, e.g., the distinction /s/ voiceless and /z/ voiced in English is under interference pressure from Spanish which has no contrast here, only /s/ with the predictable variants (s) and (z).

2. Kindergarten Achievement

Kindergarten children were tested only with the Basic Language Competence Test, sub-test No. 1, Oral Vocabulary. Children received the test individually both in Spanish and in English using different forms. They were pre-tested in January, 1973 and post-tested in June, 1973.

Table 1 reports the information related to the mean and standard deviation for the pre and post-test and also the mean gain between pre and post-test and the t-test value for the significance of this gain. The two samples were considered as independent samples, even though they were not, and, therefore, the t-test was forced somewhat.

Table 1: Oral Vocabulary: Means and Standard Deviations on Spanish and English Version of the Language Competence Test for the Kindergarten.

Language	Pre-test[1]			Post-test			Mean Gain	t-test Values
	N	Mean	SD	N	Mean	SD		
Spanish	28	15.53	7.04	46	16.23	7.76	.88	.82
English	28	21.57	3.22	46	20.71	4.09	-.86	.95

[1]Test scores are raw scores.

[2]Non-significant at the .05 level.

NOTE: Perfect score = 24.

From Table 1 we can conclude:

- The kindergarten children's Oral Vocabulary mean score was below normal level both in Spanish and English. Normal children are expected to have a perfect score of 24. Something is thwarting these children from having a perfect score and this is not normal in an American or Spanish population. In other words, we expect these children to have normal mental abilities; their mean score, therefore, is not understandable according to the standard for Spanish or American populations; they form a special kind of population by themselves both in Spanish and English.

- This group of children has a better English vocabulary than Spanish vocabulary.

- The group is a relatively homogeneous group according to their scores in English as shown by the size of the standard deviation; most of the children scored very close to the means. In Spanish, on the other hand, they form a more heterogeneous group with more variety and a greater range of scores.

- During four months in the school the children did not improve significantly in either Spanish or English. The difference between mean pre-test and mean post-test was not beyond the difference that can be expected by chance alone. If the mean gain is not significant it is useless to discuss if the gain was positive or negative. If the difference is due entirely to chance it can go up or down and, therefore, we should interpret in the same way the positive and negative mean gain if they are not significant. The t-test value was not significant beyond the .05 level.

3. First Grade Achievement

First grade children were given the Metropolitan Achievement Test, Reading, the Inter-American Series Test of Reading and the Basic Language Competence Test. In the MAT and Inter-American Series children were pre-tested at the end of kindergarten and post-tested at the end of the first grade using the same test level in both situations; in the Basic Language Competence Test the children were pre-tested in January and post-tested in June.

a. MAT-Reading

Children were tested in Listening for Sounds and Reading subtests from the MAT, Primer.

The 39 items of the Listening for Sounds subtest measure pupils' knowledge of beginning and ending sounds and sound-letter relationships. Twenty-two of these items are based on pictures; eight items are based on letters; nine items use single words.

The 33 items of the Reading subtest measure pupils' beginning reading skills. Eleven items require pupils to identify letters; fourteen items require pupils, when given a picture of some common object, to select from four words the one word that describes the picture; five items require pupils to select one of three easy sentences which best describes a picture.

1) Listening for Sounds

The information related to the Listening for Sounds subtest can be found in table 2. The information is reported in raw scores with the exception of the last column where percentiles are used.

Table 2: Listening For Sounds: Means and Standard Deviations for the Pre-test, Post-test and Gain Scores and the Correspondent Percentiles on the MAT for the First Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles[2] Pre-test	Percentiles[2] Post-test	t-test[3] Value
N	18	28				
Mean	11.20	30.35	19.15	2.4	44	8.74
Standard Deviation	6.80	7.53				

[1]The test scores are raw scores.

[2]Percentiles are for the middle of First Grade.

[3]t-test value = 8.74 which was significant beyond the .05 level.

a) Pre-program Level

From pre-test data it can be said that:

- The mean pre-test score was well below the 50th percentile of the national norms. As a matter of fact, 97.6 percent of the national population scored higher than the mean score for the children of our sample. It must be added that the national norms used were for children in the middle of the first grade while children from our sample were tested at the end of the kindergarten. This means that the national norms were developed for children a little more advanced in language skills than children in our sample and, therefore, some differences were expected.

- Children started the program being a very heterogeneous group as indicated by the size of the standard deviation. In order to include 82 percent of the population a range of about 20 points is needed; the possible range is 39 points. This means that there are scores along the whole spectrum of listening for sounds continuum, except in the highest part which was not reached by any child.

b) Pre-post Test Change

From the column labeled "Mean Gain" and the post-test data it is clear that:

- There was a positive mean gain between pre and post-testing. Students, after one year in the program, improved in their ability in Listening for Sounds.

- The improvement during one year in the program was significant beyond the .05 level. The mean gain could not be explained as a result of chance alone. Children improved in the ability measured by the subtest.

- Children at the end of the first grade were still below the 50th percentile for the national norms for children in the middle of first grade. Although children were 97.6 points below the national norms when the program started, after one year in the program, children raised their mean score to the 44th percentile, only 6 points below the national level.

- There was not a substantial change in the amount of heterogeneity of the group as indicated by the size of the standard deviations for the pre and post-test. This means that children showed similar improvements regardless of their pre-test scores. The group started as an heterogeneous group and finished with a similar degree of heterogeneity and with a wide range of scores. The program's impact on children's scores was similar throughout the whole range of scores.

2. Reading

Table 3 reports information related to the Reading subtest. The information is reported in a form similar to the previous table.

Table 3: Reading: Mean and Standard Deviation for the Pre-test, Post-test, and Gain Scores and the Correspondent Percentiles on the MAT for the First Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles[2]		[3] t-test Value
				Pre-test	Post-test	
N	20	28				
Mean	9.20	27.71	18.51	4	85	44.35
Standard Deviation	4.20	5.73				

[1]The test scores are raw scores.

[2]Percentiles are for the middle of First Grade.

[3]t-test value = 44.35 that was significant beyond the .05 level.

a) Pre-program Level

From this pre-test data, both the raw scores and the percentiles, it can be concluded that:

- The mean pre-test scores were well below the national norms. Students as a group showed less achievement in Reading than the mean for a national sample. Ninety-six percent of the students of tne national population scored higher than the mean score for our

sample. It must be noted again that the national norms are for children in the middle of first grade.

- Children started the program being a heterogeneous group as indicated by the size of the standard deviation. In order to include 82 percent of the population, a range of about 13 points is needed; the possible range is 33 points. Only 2 percent of the sample however, had a score greater than 20.

b) Pre-post Test Change

From the column labeled "Mean Gain" and the post-test data it can be said that:

- There was a positive mean gain between pre and post-testing. Students, after one year in the program, improved in their ability in Reading.

- The improvement was significant beyond the .05 level. The mean gain could not be explained as a result of chance alone. Children improved in the ability measured by the subtest.

- Children at the end of the first grade were well above the 50th percentile for the national norms for children in the middle of first grade. Children, after one year in the program, raised their mean reading score to the 85th percentile, 35 points above the national level and 74 points above their mean percentile when the program started. The program's impact in reading has been substantial.

- There was little change in the heterogeneity of the group as indicated by the size of the standard deviations for the pre and post-test. Children's scores at the end of the program had a greater range than at the beginning of the program. Children showed slight dissimilarities in improvement during a year in the program.

c. Comparison of Subtests

Table 4 contains the information related to each subtest for the first grade. The following are conclusions from the analysis of this table:

Table 4: Mean and Standard Deviations for the Pre-test, Post-test and Percentiles on Each Subtest of the MAT-Reading for First Grade.

Test[3]	Pre-test[1]			Post-test			Mean Gain	Percentiles[4]		t-test Values
	N	Mean	SD	N	Mean	SD		Pre-test	Post-test	
Listening For Sounds	18	11.2	6.80	28	30.35	7.53	19.15	2.4	44	8.74*[2]
Reading	20	9.2	4.20	28	27.71	5.29	18.51	4.0	85	44.35*

[1]The test scores are raw scores.

[2]*t-test value significant beyond the .05 level.

[3]Correlation between Listening for Sounds and Reading : r=.30.

[4]Percentiles are for the middle of first grade.

- Students had almost the same mean score in each subtest in relation to the national norms when the program started. The percentile for the mean scores in Listening for Sounds and in Reading was also close with a difference of 1.6 percentage points.

- Students scored substantially lower than the national sample in both Listening for Sounds and in Reading at the beginning of the program. Their mean score was 48 and 46 percentiles below the mean for the national norms , respectively.

- Students showed a positive and significant improvement between pre and post-testing in both subtests. In both cases the t-test values for the gains were large enough to make the gains improbable as a result of chance alone.

- The mean raw score gain was very similar in both subtests but there were striking differences in percentiles. Students showed an improvement of 41 percentile points in Listening for Sounds while their improvement in Reading was 81 percentile points.

- Students at the end of the first grade were slightly below the national norms for middle of first grade in Listening for Sounds but they were well above the national norms in Reading.

- The correlation between raw scores in both subtests was low, the product-moment correlation being .30. This coefficient means that not more than 9 percent of the scores in one subtest can be predicted from or is related to the scores in the other subtest. In practical terms the finding indicates that scores on Listening for Sounds and Reading measure different traits. Even though both traits are related as part of language skills, nevertheless, they show sufficient empirical independence not to be treated in the same way or as a package.

b. Inter-American Series Test of Reading

The Test of Reading in English was given to all children, either Spanish or English dominant. The Test of Reading in Spanish was given only to those students with some ability in Spanish. Level I was used with first grade children. Form CE in English and in Spanish was used for the post-test.

The test was developed for children in grades 1 to 6. Our sample was in kindergarten at the time of pre-testing and, therefore, we do not have pre-test data.

Caution should be indicated in attributing too much significance to percentile ranks as this test was normed on small, specific groups of children. These groups were, in addition, only tested in their dominant language whereas the sample at C.E.S. 235 was tested in their secondary language as well.

The Vocabulary subtest measures the child's ability to match pictures with words. The Comprehension subtest measures the student's ability to relate sentences and stories to pictures. The total is the sum of the two subtests.

1) Test of Reading in Spanish

Table 5 has the information about the means and standard deviations and percentile ranks for the post-test scores. Children were not pre-tested and so no information about progress during one year in the program is available. From this table we have the following important information:

Table 5: Means and Standard Deviations for the Post-test, and Percentile Ranks of Each Subtest of the Inter-American Series Test of Reading, in Spanish, Level I, for the First Grade.

Test[3]	N	Mean[1]	SD	Percentiles[2]
Vocabulary	14	13.64	9.22	42
Comprehension	14	10.21	6.32	35
Total Reading	14	23.85	14.79	40

[1]Test scores are raw scores.

[2]Percentiles are for the First Grade from Puerto Rico.

[3]Correlation between subtests: r = .80.

- Children in school C.E.S. 235 scored below the norms for children of first grade in Puerto Rico. The mean percentile ranks for our sample, in Vocabulary, Comprehension and Total Reading, was below the 50th percentile of first grade children from Puerto Rico. Children in our sample had less knowledge of Spanish words and less ability to understand written statements in Spanish than children of their age in Puerto Rico. The findings was expected because of cultural and language differences between Puerto Rican children living on the island and Spanish speaking children living in the United States.

- Children in our sample scored slightly lower in Comprehension than in Vocabulary. The difference between the two subtests was only 7 percentile points. The children had more difficulty in relation to the norms in understanding written statements than in knowledge of words.

- There was a great variety of scores both in Vocabulary and Comprehension within the group as indicated by the size of the standard deviations which can be considered large. In order to include 96 percent of the scores a range of 36 points is needed in Vocabulary and 25 in Comprehension. The perfect score in both subtests is 40.

- The correlation between Vocabulary and Comprehension was really high and similar to the coefficient found among other Spanish groups as reported in the Technical Report of the Inter-American Series. This coefficient means that about 64 percent of the differences in scores in Comprehension are explained by differences in Vocabulary. In order to understand a written statement in Spanish a knowledge of Vocabulary is very important, and progress in one of the subtests is followed by very similar progress in the other subtest.

2) Test of Reading in English

The information in relation to the English version of the Test of Reading is reported in Table 6 in a form similar to the form used in the previous table for the Spanish version of the same test. From this table we may conclude that:

Table 6: Means and Standard Deviations for the Post-test, and Percentile Ranks of Each Subtest of the Inter-American Series Test of Reading, in English, Level I, for the First Grade.

Test[3]	N	Mean	SD	Percentiles[2]
Vocabulary	28	21.92	7.95	28
Comprehension	28	16.32	9.61	27
Total Reading	28	38.25	16.45	25

[1]Test scores are raw scores.

[2]Percentiles are for the First Grade.

[3]Correlation between Vocabulary & Comprehension: r=.75

- Children in school C.E.S. 235 scored below the norms for American children in first grade. The mean percentile ranks in Vocabulary and Comprehension were well below the 50th percentile of

American norms. Children in our sample had less knowledge of English words and less ability to understand written statements in English than American children. The finding was expected because of cultural and language differences between American children and Spanish children living in the USA, as most of the students in C.E.S. 235 are.

- The mean percentile ranks were very similar in both subtests; the difference was only 1 percentile point. Children in our sample scored behind American norms both in Vocabulary and Comprehension with almost identical ranking in both subtests.

- There was a great variety of scores both in Vocabulary and Comprehension within the group, as indicated by the size of the standard deviation which can be considered as large. In order to include 96 percent of the scores a range of about 32 points is needed in Vocabulary and about 38 points in Comprehension. The perfect score in both subtests is 40.

- The correlation between Vocabulary and Comprehension was high and similar to the coefficient found among American groups as reported in the Technical Report. This coefficient means that about 56 percent of the differences in scores are explained by differences in Vocabulary. In order to understand a written statement in English, a knowledge of vocabulary is very important, and progress in one of the subtests is followed by similar progress in the other subtest.

c. Tests of Basic Language Competence

Tests of Language Competence in Spanish and English were given to all children who were judged by the teachers to be minimally competent in Spanish or English. Only subtests 1 and 2 were given. A different form was used in Spanish and in English. Children were pre-tested in January and post-tested in June, 1973.

Since these tests have not been normed on a national sample, percentile ranks are not included here. The only possible comparison is to a perfect total score which is reported at the bottom of each table.

The test of Language Competence was designed to measure a child's ability to understand Spanish and English. Children who speak one or both languages as a native language are expected to make near perfect scores on the subtests.

The subtest, Oral Vocabulary, presents words to the child from which the child must identify the corresponding picture. Table 7 has the pertinent information on this subtest. From the data on this table we can conclude that:

Table 7: Oral Vocabulary Means and Standard Deviations on Spanish and English Version of the Basic Language Competence Test for the First Grade.

Language	Pre-test[1]			Post-test			Mean Gain	t-test[2] Values
	N	Mean	SD	N	Mean	SD		
Spanish	40	15.20	7.49	38	14.55	8.63	-.65	.34
English	40	20.85	3.86	38	22.76	1.45	1.91	1.50

[1]Test scores are raw scores.

[2]Non-significant at the .05 level.

NOTE: Perfect score = 24.

- Children in school C.E.S. 235 scored relatively low both in Spanish and in English in the pre-test. Children started the program with a knowledge of English and Spanish words that was less than perfect.

- Children started the program with a better knowledge of English than Spanish words. Their mean score in English was closer to a perfect mean score than the mean score in Spanish.

- The children formed relatively heterogeneous groups according to their scores in Spanish and in English, and they were more heterogeneous in Spanish than in English as indicated by the large size of the standard deviations. In order to include 96 percent of the sample children we need a complete range of all the possible scores in Spanish. This means that some of the children had a perfect score and others were near zero score.

- Children did not show any significant change in Oral Vocabulary during five months of school in either Spanish or English. The mean gain between pre and post-test scores was not large enough to be significant at the .05 level as indicated by the t-test values. The mean gains both in Spanish and in English can be explained as a result of chance alone and they do not represent real changes in vocabulary.

- A significant finding was that children became a more homogeneous group in English. There was a substantial reduction in the size of standard deviation. There was a small increase in the group's heterogeneity in Spanish.

4. Second Grade Achievement

Second grade children were given the Metropolitan Achievement Test, Reading, the Inter-American Series Test of Reading and the Basic Language Competence Test. Children were pre-tested at the end of the first grade and post-tested at the end of the second grade using the same test level in both situations.

a. MAT-Reading

Children were tested in Word Knowledge, Word Analysis and Reading subtests from the MAT, Level Primary I.

The 35 items of the Word Knowledge subtest measure the extent of a pupils' reading vocabulary. Pupils are given a picture of some common object and must select from four words the one word that describes the picture. Words are generally from primary level readers.

The 40 items of the Word Analysis subtest measure pupils' knowledge of sound-letter relationships or skill in decoding. Pupils must identify a dictated word from among several words with similar configurations and sound patterns.

The 42 items of the Reading subtest measure pupils' comprehension of written material. Thirteen items require pupils to select one of three easy sentences which best describes a picture. Nineteen items require pupils to read simple paragraphs and answer questions about what they have read.

1) Word Knowledge

The information related to the Word Knowledge subtest is reported in Table 8. This table contains information about the number of students and mean grade equivalent on the pre-test, post-test, gain between pre and post-test and difference between predicted and actual post-test.

Table 8: Word Knowledge: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Second Grade.

	Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2]	
						Mean Gain	Difference
N	32	31					
Mean	1.60	2.61	1.01	2.43	.18	4.83*	.56
Standard Deviation	.62	1.00		1.47			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 1.80 at the time of the pre-test and 2.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than was expected of them if they conformed to the national sample. The result, though, was expected given the special characteristics of school C.E.S. 235.
- The mean grade equivalent was 2 months below the national level. This means that children in the program finished first grade with achievement scores slightly lower than the national norms.
- There were differences among the students in their Word Knowledge scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample, a range of 1 year, 2 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

The column labeled "Mean Gain" merely indicates the difference between the pre-test and the post-test mean scores. From this column we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 10 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value, and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 1 month during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 1 month lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished one year of school being a heterogeneous group, and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

The column labeled "Predicted Post-test" indicates a statistical calculation based on pre-test scores. Each student's achievement thus far in school is divided by the number of months he has been in school (excepting kindergarten) and from this monthly mean gain the predicted post-test scores can be determined by simply multiplying the monthly mean gain by the number of months in the school between pre and post-testing and adding the result to the pre-test scores. The reason for this statistical manipulation is to determine whether the children gain more during the period in the program than they would have if they had continued making progress at the same rate they have thus far in their school career.

It should be noted that the predicted post-test scores are a direct reflection of the pre-test scores. If pre-test scores are low, a sample has gained very little up to that point and would be expected to gain very little in the future without some special treatment. None of the school samples were expected to gain to their grade level if they were not at that level when they were pre-tested.

The column labeled "Difference" is the most important of the chart. It indicates the difference between what the children were expected to achieve and what they did achieve. From that column, the following conclusions can be drawn:

- The difference between actual and predicted post-test scores was positive. Children's mean scores at the end of the year were higher than the mean expectations based on their scores at the end of the previous year. The school's impact was beneficial for the students as a group; during the year, students improved at a rate which was greater than the monthly rate of improvement the previous year.

- The mean difference between predicted and actual post-test was not significant at the .05 level as indicated by the t-test value. The positive improvement of about 2 months beyond expectations that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the improvement was a real improvement in the ability measured by the test, it could be explained by chance alone.

2) Word Analysis

All the information about the Word Analysis subtest is reported in Table 9. The information is reported in a form that is similar to the form used in the previous table.

Table 9: Word Analysis: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Second Grade.

	Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2]	
						Mean Gain	Difference
N	29	31					
Mean	1.60	2.32	.72	2.43	-.11	4.40*	.43
Standard Deviation	.51	.73		1.21			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 1.80 at the time of the pre-test and 2.90 at the time of the post-test.

a) Pre-program Level

The data from the pre-test supports the following conclusions:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, again, was expected given the special characteristics of school C.E.S. 235.
- The mean grade equivalent was 2 months below the national level. This means that children in the program finished first grade with achievement scores slightly lower than the national norms.
- There were differences among the students in their Word Analysis scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample, a range of 1 year above and below the mean would be necessary. The sample was a heterogeneous group with a relatively wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "Mean Gain", we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 7 months higher than their mean score a year earlier. The mean gain was significant beyond the .05 level and it could not be explained by chance alone; it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 4 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school, the students were 4 months lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviation for the pre and post-test. Students, as a group, finished one year of school being a heterogeneous group and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

From the column labeled "Difference" we can conclude that:

- The difference between actual and predicted post-test scores was negative. Children's mean score at the end of the year was lower than the mean expectations based on their scores at the end of the previous year. The school's impact on students, as a group, was less than expected. During the year, students improved at a rate which was smaller than the monthly rate of growth the previous year.

- The mean difference between predicted and actual post-test, however, was not significant at the .05 level as indicated by the t-test value. The negative mean change of about 1 month short of expectations was not large enough to be significant beyond changes expected by chance alone. We can not say that the negative change was a real change in the ability measured by the test.

3. Reading

The information about the Reading subtest is reported in Table 10 in a form which is similar to the form used in the previous tables.

Table 10: Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Second Grade.

	Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2]	
						Mean Gain	Difference
N	31	31					
Mean	1.40	2.50	1.10	1.95	.55	5.09*	3.13*
Standard Deviation	.38	.95		.90			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 1.80 at the time of the pre-test and 2.90 at the time of the post-test.

a) Pre-program Level

From the data for the pre-test scores as reported in the table, it can be said that:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, again, was expected given the special characteristics of school C.E.S. 235.

- The mean grade equivalent was 4 months below the national level. This means that children in the program finished first grade with achievement scores lower than the national norms.

- There were small differences among the students in their Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 8 months above and below the mean would be necessary. The sample was a heterogeneous group .

b) Pre-post Test Change

From the column labeled "Mean Gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 11 months higher than their mean score a year earlier. The mean gain was significant beyond the .05 level; it could not be explained by chance alone: it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level remained the same during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school, the students were at the same level in relation to the national sample as they were a year earlier. Students improved during the year as much as the national sample but not more and, therefore, the gap between mean grade equivalent and national level at the end of the year was the same as the previous year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviation for the pre and post-test. Students, as a group, finished one year of school being a heterogeneous group and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

From the column labeled "Difference" it can be said that:

- The difference between actual and predicted post-test scores was positive. Children's mean scores at the end of the year was higher than the mean expectation based on their scores at the end of the previous year. The school's impact was beneficial for the students as a group; during the year students improved at a rate which was greater than the monthly rate of improvement the previous year.

- The mean difference between predicted and actual post-test was significant beyond the .05 level as indicated by the t-test value. The positive improvement of about 5 and 1/2 months beyond expectations that occurred in the students as a sample was large enough to be significant beyond the changes expected by chance alone. We can say that the improvement was a real improvement in the ability measured by the test and it can not be explained by chance alone.

4) Total Reading

The information about Total Reading is reported in Table 11 in a form similar to the form used in the previous tables. Total Reading scores are simply the sum of Word Knowledge and Reading scores.

Table 11: Total Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Second Grade.

	Pre Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2] Mean Gain	Difference
N	31	31					
Mean	1.50	2.45	.95	2.19	.26	5.21*	1.25
Standard Deviation	.36	.78		.86			

[1] The test scores are grade equivalents.

[2] *Significant beyond the .05 level.

NOTE: The National Grade Level was 1.80 at the time of the pre-test and 2.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, though, was expected given the special characteristics of school C.F.S. 235.

- The mean grade equivalent was 3 months below the national level. This means that children in the program finished first grade with achievement scores slightly lower than the national norms.

- There were differences among the students in their Total Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 8 months above and below the mean would be necessary. The sample was a heterogeneous group.

b) Pre-post Test Change

From the column labeled "Mean Gain", we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 9 and 1/2 months higher than their mean score a year earlier. The mean gain was significant beyond the .05 level; it could not be explained by chance alone; it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 1 and 1/2 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 1 and 1/2 month lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished one year of school being a heterogeneous group and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

From the column labeled "Difference" we can make the following conclusions:

- The difference between actual and predicted post-test scores was positive. Children's mean scores at the end of the year was higher than the mean expectations based on on their scores at the end of the previous year. The school's impact was beneficial for the students as a group; during the year students improved at a rate which was greater than the monthly rate of improvement the previous year.

- The mean difference between predicted and actual post-test was not significant at the .05 level as indicated by the t-test value. The positive improvement of about 2 and 1/2 months beyond expectations that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the improvement was a real improvement in the ability measured by the test and it can be explained by chance alone.

5) Comparison of Subtests

Table 12 includes information about each subtest and the total test. Information is arranged in a form similar to the form used in the previous tables. From this table we have the following additional information:

Table 12: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test and t-test Values for the Difference Between Pre-test, Actual Post-test and Predicted Post-test on Each Subtest of the MAT-Reading for the Second Grade.

Test	Pre-test[1]			Actual Post-test			Mean Gain	Predicted Post-test		Difference	t-test Values	
	N	Mean	SD	N	Mean	SD		Mean	SD		2	3
Word Knowledge	32	1.60	.62	31	2.61	1.00	1.01	2.43	1.47	.18	4.83*	.55
Word Analysis	29	1.60	.51	31	2.32	.73	.72	2.43	1.21	-.11	4.40*	.43
Reading	31	1.40	.38	31	2.50	.95	1.10	1.95	.90	.55	5.09*	3.13*
Total Reading	31	1.50	.36	31	2.45	.78	.95	2.19	.86	.26	5.21*	1.25

[1]The test scores are grade equivalent.

[2]t-test values for the difference between Pre and Actual Post-test Mean Scores.

[3]t-test values for the difference between Actual and Predicted Post-test Mean Scores.

*t-test value significant beyond the .05 level.

NOTE: The National Grade Level was 1.80 at the time of the pre-test and 2.90 at the time of the post-test.

a) Pre-program Level

From pre-test scores it can be said that:

- Children scored below national norms in the three subtests, Word Knowledge, Word Analysis and Reading.
- Children finished first grade 2 months below national norms in Word Knowledge and Word Analysis and 4 months below national norms in Reading.
- The gap between grade equivalent and national norms was very similar in every subtest; students scored 2 months lower in Reading than in Word Knowledge and Word Analysis.
- Students were heterogeneous groups in every subtest according to their scores. The degree of heterogeneity was different in every subtest; the greatest degree of heterogeneity occurred in Word Knowledge and the smallest in Reading.

b) Pre-post Test Change

From the column labeled "Mean Gain " the following conclusions can be drawn:

- There was a positive gain in grade equivalent between pre and post-test. Children, after eleven months in school, scored higher than when they finished the year before. They improved in all the subtests.

- Children improved less in Word Analysis than they did in Word Knowledge and Reading, the difference between mean gains being 3 and 4 months respectively.

- The gap between mean grade equivalent and the national norms increased in Word Knowledge and Word Analysis but it remained the same in Reading. During eleven months of school, children in the sample improved 11 months in Reading but they improved only 10 months in Word Knowledge and 7 months in Word Analysis, thus increasing the gap by 1 month in Word Knowledge and 4 months in Word Analysis.

- There were substantial increases in group heterogeneity in every subtest as indicated by the size of the standard deviations for the pre and post-test.

c) Program Impact

From the column labeled "Difference" it can be concluded that:

- There was a positive difference in Word Knowledge and Reading between predicted and actual post-test, but the difference was negative in Word Analysis.

- The program's impact was beneficial both in Word Knowledge and Reading and on the Total Reading. Students scored higher than was expected of them, with expectation based on their past improvement.

- The program's impact was not beneficial in Word Analysis. Children scored lower

than was expected of them according to their scores the previous year.

- Change beyond expectations was significant only in Reading. The changes in Word Knowledge, Word Analysis and Total Reading beyond expectations can be explained as a result of chance alone.

d) Correlation Between Subtests

Table 13 reports the correlations between the three subtests for school CES 235 and for a national sample.

Table 13: Product-Moment Correlation Coefficients Between Word Knowledge, Word Analysis and Reading from the MAT for the Second Grade Children from School CES 235 and from a National Sample.

	C.E.S. 235	National Sample
Word Knowledge Vs. Word Analysis	.70	.88
Word Knowledge Vs. Reading	.65	.80
Word Analysis Vs. Reading	.50	.79

From this table we have the following conclusions:

- The correlations between subtests were moderate in school CES 235. This means that changes in scores in a subtest are followed by similar changes in the other subtests. The amount of change in a subtest that can be explained from the changes in the other is 25 percent between Word Analysis and Reading, 49 percent between Word Knowledge and Word Analysis and about 42 percent between Word Knowledge and Reading.

- The coefficients are substantially lower in school CES 235 than in the national sample. This result is in accordance with what has been demonstrated when a test is used with groups of different characteristics than the groups for which it was developed; the coefficients usually go down. What this finding means is that in school CES 235 children improved less in a

subtest as result of improvements on other subtests than in the national sample. The different reading skills are more interrelated in a national sample than in school CES 235 where children's improvements in the various reading skills are relatively more independent.

b. Inter-American Series Test of Reading

The Test of Reading in English was given to all children, either Spanish or English dominant. The Test of Reading in Spanish was given only to those students with some ability in Spanish. Level I was used with second grade children. Form CE in English and in Spanish was used for the pre-test. In the post-test form CE was used in English and form DEs in Spanish.

Caution should be used against attributing too much significance to percentile ranks as this test was normed on small, specific groups of children. These groups were, in addition, only tested in their dominant language whereas the sample at C.E.S. 235 was tested in their Secondary language as well.

The Vocabulary subtest measures the child's ability to match pictures with words. The Comprehension subtest measures the student's ability to relate sentences and stories to pictures. The total is the sum of the two subtests.

1) Test of Reading in Spanish

a) Vocabulary

Table 14 contains the information about the means and standard deviations for the pre and post-test, the mean gain between pre and post-test, the percentile ranks and the t-test value for the mean gain.

Table 14: Vocabulary: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in Spanish, Level I for the Second Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles[2]			t-test[3] Values
				Pre-test	Post-test	Gain	
N	36	22					
Mean	15.80	27.18	11.38	50	68	18	3.76
Standard Deviation	11.20	11.13					

[1] Test scores are raw scores.

[2] Percentiles are for First Grade from Puerto Rico.

[3] t-test value = 3.76 which was significant beyond the .05 level.

From this table we have the following information:

- Children in school CES 235 finished the first grade with a mean score in Spanish Vocabulary that was exactly the 50th percentile of the norms for first grade children in Puerto Rico. This means that 50 percent of children in Puerto Rico scored below the mean vocabulary score for children in our sample. This result was not expected; cultural and language differences between Puerto Rican children living on the island and Spanish speaking children living in the United States were expected to be reflected in vocabulary scores but they were not.

- There was a great variety of scores in Spanish Vocabulary within the group as indicated by the size of the standard deviation for the pre-test. In order to include 96 percent of the scores from the sample a range of 44 points would be needed; the possible range is 40 points. This apparent contradiction is explained by the fact that the sample can be divided in two group of scores; some of the children had an almost perfect score, and some had a very low score with

very few of the children scoring between these two extremes.

- During eleven months of school, children improved in their Spanish Vocabulary scores. The mean score increased by 18 percentile points. There are no norms for second grade children in Puerto Rico and, therefore, the only comparison we can make is with the first grade. The findings means that second grade children from C.E.S. 235 had a mean score which was 18 percentile points higher than the 50th percentile for the first grade children in Puerto Rico and also 18 percentile points higher than their own mean score at the end of the first grade.

- The mean gain of 18 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in Spanish Vocabulary was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was not any substantial change in group variability between pre and posttesting as indicated by the size of the standard deviations for the pre and posttest. Children finished the first grade being a heterogeneous group according to their scores in Spanish Vocabulary and they finished the second grade with almost the same degree of heterogeneity.

b) Comprehension

The information related to the Spanish Comprehension subtest is reported in Table 15 in a form that is similar to the form used in the previous tables. The following information is drawn from this table:

Table 15: Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in Spanish, Level I for the Second Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles[2]			t-test[3] Values
				Pre-test	Post-test	Gain	
N	25	22					
Mean	14.30	20.72	6.42	57	80	23	2.43
Standard Deviation	8.40	9.71					

[1] Test scores are raw scores.

[2] Percentiles are for First Grade from Puerto Rico.

[3] t-test value = 2.43 which was significant beyond the .05 level.

- Children in school C.E.S. 235 finished first grade with a mean Spanish Comprehension score higher than the 50th percentile of the first grade children in Puerto Rico. Children from our sample had a mean reading comprehension in Spanish greater than 50 percent of the scores for children of the same age in Puerto Rico. This finding, again, was not expected because of the particular characteristics of school CES 235 children.

- There was a great variety of scores in Spanish Comprehension within the group as indicated by the size of the standard deviation for the pre-test. In order to include 96 percent of the scores from the sample a range of about 34 points would be needed when the possible range is 40.

- During eleven months of school, children improved in their Spanish Comprehension scores. The mean score increased by 23 percentile points. It must be added that the percentile rank norms are for first grade children while children in our sample were at the end of the second grade. The finding means that second grade childrens from CES 235 had a mean score which was 30 percentile points higher than the 50th percentile for the first grade child-

ren in Puerto Rico and also 23 percentile points higher than their own mean score at the end of the first grade.

- The mean gain of 23 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in Spanish Comprehension was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a small increase in group variability between pre and post-testing as indicated by the size of the standard deviations for the pre and post-test. Children in school CES 235 finished first grade being a heterogeneous group according to their scores in Spanish Comprehension and they finished second grade with a greater degree of heterogeneity.

c) Total Reading

The information related to the Spanish Total Reading subtest is reported in Table 16 in a form that is similar to the form used in the previous tables. The following information is drawn from this table:

Table 16: Total Reading: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Sereis Test of Reading, in Spanish, Level I for the Second Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles[2]			t-test[3] Values
				Pre-test	Post-test	Gain	
N	25	22					
Mean	32.10	47.90	15.80	56	73	17	6.88
Standard Deviation	18.40	19.26					

[1]Test scores are raw scores.

[2]Percentiles are for First Grade from Puerto Rico.

[3]t-test value = 6.88 which was significant beyond the .05 level.

- Children in school CES 235 finished first grade with a mean Spanish Total Reading score higher than the 50th percentile of the first grade children in Puerto Rico. Children from our sample had a mean Total Reading in Spanish greater than 50 percent of the scores for children of the same age in Puerto Rico. This finding, again, was not expected because of the particular characteristics of school CES 235 children.

- There was a great variety of scores in Spanish Total Reading within the group as indicated by the size of the standard deviation for the pre-test. In order to include 96 percent of the scores from the sample a range of about 74 points would be needed when the possible range is 80.

- During elven months of school, children improved in their Spanish Total Reading scores. The mean score increased by 17 percentile points. It must be added that the percentile rank norms are for first grade children while children in our sample were at the end of the second grade. The finding means that second grade children from CES 235 had a means score which was 23 percentile points higher than the 50th percentile for the first grade children in Puerto Rico and also 17 percentile points higher than their own mean score at the end of the first grade.

- The mean gain of 17 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in Spanish Total Reading was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a small increase in group variability between pre and post-testing as indicated by the size of the standard deviations for the pre and post-test. Children in school CES 235 finished first grade being a heterogeneous group according to their scores in Spanish Total Reading and they finished second grade with a greater degree of heterogeneity.

d) Comparison of Subtests

Table 17 includes information about each subtest and the total test. Information is arranged in a form similar to the form used in the previous tables. From this table we have the following additional information:

Table 17: Means and Standard Deviations for the Pre-test and Post-test, Percentile Ranks of each Subtest of the Inter-American Series Test of Reading, in Spanish, Level I, for the Second Grade.

Test[4]	Pre-test[1]			Post-test			Mean Gain	Percentiles[2]		t-test[3] Values
	N	Mean	SD	N	Mean	SD		Pre-test	Post-test	
Vocabulary	36	15.80	11.20	22	27.18	11.13	11.38	50	68	3.76*
Comprehension	25	14.30	8.40	22	20.72	9.71	6.42	57	80	2.43*
Total Reading	25	32.10	18.40	22	47.90	19.26	15.80	56	73	6.88*

[1]Test scores are raw scores.

[2]Percentiles are for First Grade from Puerto Rico.

[3]*Significant beyond the .05 level.

[4]Correlation between subtests: r= .70.

- Children in school CES 235 finished first grade with a mean score in Spanish Vocabulary, Comprehension and Total Reading as high or higher than the 50th percentile of the first grade children in Puerto Rico.

- There was a great variety of scores both in Spanish Vocabulary and Comprehension as indicated by the size of the standard deviations for the pre-test. Children in school CES 235 formed a more heterogeneous group in Vocabulary than in Comprehension although there was a wide range of scores in both subtests.

- During eleven months of school, children improved in both subtests scores. They improved more in Comprehension than in Vocabulary. The difference between pre

and post-test was 18 percentile points in Vocabulary and 23 points in Comprehension.

- The mean gains in both subtests and Total Reading were significant beyond the .05 level as indicated the the t-test values. The improvements were greater than what could be expected by chance alone, and they represented real improvements in the abilities measured by the subtests.

- There was either no change or a small increase in group variability between pre and post-testing in the subtests. Children in school CES 235 finished first grade being a heterogeneous group according to their scores in Spanish Vocabulary and Comprehension and they finished second grade with the same or greater degree of heterogeneity.

- The correlation between raw scores in Spanish Vocabulary and Comprehension was moderate. A coefficient of .70 means that 49 percent of the scores in one of the subtests can be predicted from the scores in the other subtest. Changes in one of the subtest will be followed by similar changes, until a certain degree, in the other subtest. The coefficient, however, is smaller than the coefficients reported in the manual which are in the 80's.

2) Test of Reading in English

a) Vocabulary

Table 18 contains the information about the English Vocabulary subtest of the Reading Test in English. The information is reported in a form similar to the form used in the Spanish version of the same test. The following information is drawn from this table:

Table 18: Vocabulary: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English, Level I, for the Second Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles[2]			t-test[3] Values
				Pre-test	Post-test	Gain	
N	29	30					
Mean	16.10	28.36	12.26	15	52	37	4.19
Standard Deviation	10.80	11.63					

[1]Test scores are raw scores.

[2]Percentiles are for First Grade.

[3]t-test value = 4.19 which was significant beyond the .05 level.

- Children in school CES 235 finished first grade with a mean score in English Vocabulary that was well below the 50th percentile for native English speaking first grade children. About 85 percent of first grade children from a national sample scored higher than the mean score of the children in our sample. The result was expected because of cultural and language differences between native English speaking children and children in school CES 235.

- There was a great variety of scores in English Vocabulary within the group as indicated by the size of the standard deviations for the pre-test. A range of 43 points would be needed to include 96 percent of the scores from the sample; the possible range is 40 points. The sample could be divided in two separate groups according to their scores; some of the children had an almost perfect score, and some had a near zero score with very few scores in between. This explains the apparent contradiction between a possible range of 40 and a needed range of 43 in the sample.

- During eleven months of school, children improved in their English Vocabulary scores. The mean score increased by 37 percentile points. We must add that our sample children were post-tested at the end of their second grade and the norms are for children in first grade. The finding means that second grade children in school CES 235 had a mean knowledge of English Vocabulary which was only 2 percentile points above the 50th percentile of the norms for first grade children, and it was 37 percentile points higher than their mean score a year before.

- The mean gain of 37 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in English Vocabulary was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a small increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores in English Vocabulary at the end of first grade and they had a similar degree of heterogeneity at the end of second grade.

b) Comprehension

Table 19 has the information related to the Comprehension subtest. The information is reported in a form similar to the form used in previous tables. From this table we draw the following conclusions:

Table 19: Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English, Level I, for the Second Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles[2]			t-test[3] Values
				Pre-test	Post-test	Gain	
N	28	30					
Mean	14.50	27.56	13.06	22	68	46	4.52
Standard Deviation	11.20	10.80					

[1]Test scores are raw scores.

[2]Percentiles are for First Grade.

[3]t-test value = 4.52 which was significant beyond the .05 level.

- Children in school CES 235 finished first grade with a mean score in English Comprehension that was well below the 50th percentile for native English speaking first grade children. About 78 percent of first grade children from a national sample scored higher than the mean score of the children in our sample. The result was expected because of cultural and language differences between native English speaking children and children in school CES 235.

- There was a great variety of scores in English Comprehension within the group as indicated by the size of the standard deviation for the pre-test. A range of 45 points would be needed to include 96 percent of the scores from the sample; the possible range is 40 points. The sample could be divided in two separate groups according to their scores; some of the children had an almost perfect score, and some had a near zero score with very few scores in between. This explains the apparent contradiction between a possible range of 40 and a needed range of 45 in the sample.

- During eleven months of school, children improved in their English Comprehension scores. The mean score increased by 46 points. We must add that our sample children were post-tested at the end of their second grade and the norms are for children in first grade. The finding means that second grade children in school CES 235 had a mean knowledge of English Comprehension which was 18 percentile points above the 50th percentile of the norms for first grade children and it was 46 percentile points higher than their mean score a year before.

- The mean gain of 46 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in English Comprehension was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a small decrease in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores in English Comprehension at the end of first grade and they had a similar degree of heterogeneity at the end of second grade.

c) Total Reading

Table 20 contains the information related to English Total Reading in a form that is similar to the form used in previous tables. From data in this table we can draw the following conclusions:

Table 20: Total Reading: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English, Level I for the Second Grade.

	[1] Pre-test	Post-test	Mean Gain	Percentiles[2] Pre-test	Post-test	Gain	[3] t-test Values
N	28	30					
Mean	30.00	55.93	25.93	16	62	46	4.14
Standard Deviation	21.30	22.14					

[1] Test scores are raw scores.

[2] Percentiles are for First Grade.

[3] t-test value = 4.14 which was significant beyond the .05 level.

- Children in school CES 235 finished first grade with a mean score in English Total Reading that was well below the 50th percentile for native English speaking first grade children. About 84 percent of first grade children from a national sample scored higher than the mean score of the children in our sample. The result was expected because of cultural and language differences between native English speaking children and children in school CES 235.

- There was a great variety of scores in English Total Reading within the group as indicated by the size of the standard deviation for the pre-test. A range of 85 points would be needed to include 96 percent of the scores from the sample; the possible range is 80 points. The sample could be divided in two separate groups according to their scores; some of the children had an almost perfect score, and some had a near zero score with very few scores in between. This explains the apparent contradiction between a possible range of 80 and a needed range of 85 in the sample.

- During eleven months of school, children improved in their English Total Reading scores. The mean score increased by 46 points. We must add that our sample children were post-tested at the end of their second grade and the norms are for children in first grade. The finding means that second grade children in school CES 235 had a mean knowledge of English reading which was 12 percentile points above the 50th percentile of the norms for first grade children, and it was 46 percentile points higher than their mean score a year before.

- The mean gain of 46 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in English Total Reading was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a small increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores in English Total Reading at the end of first grade and they had a similar degree of

heterogeneity at the end of second grade.

d) Comparison of Subtests

Table 21 includes the information related to each subtest and the total test. From this table we have the following additional information:

Table 21: Means and Standard Deviations for the Pre-test and Post-test, Percentile Ranks of each Subtest of the Inter-American Series Test of Reading, in English, Level I for the Second Grade.

Test[4]	Pre-test[1]			Post-test			Mean Gain	Percentiles[2]		t-test[3] Value
	N	Mean	SD	N	Mean	SD		Pre-test	Post-test	
Vocabulary	29	16.10	10.80	30	28.36	11.63	12.26	15	52	4.19*
Comprehension	28	14.50	11.20	30	27.56	10.80	13.06	22	68	4.52*
Total Reading	28	30.00	21.30	30	55.93	22.14	25.93	16	62	4.14*

[1]Test scores are raw scores.

[2]Percentiles are for First Grade.

[3]*Significant beyond the .05 level.

[4]Correlation between subtests: r = .94.

- Children in school CES 235 finished first grade , as reported by the pre-test data, with a mean score in English Vocabulary, Comprehension and Total Reading well below the 50th percentile of the native English speaking first grade children.

- There was a great variety of scores in both subtests, English Vocabulary and Comprehension, as indicated by the size of the standard deviations for the pre-test. The degree of heterogeneity was similar in both subtests.

- During eleven months of school, children improved in both subtests scores; however, they improved more in Comprehension than in Vocabulary. The difference between pre and post-test was 37 percentile points in Vocabulary and 46 in Comprehension.

- The mean gain in both subtests and Total Reading were significant beyond the .05 level as indicated by the t-test values. The improvements were greater than what could be expected by chance alone and they represented real improvements in the abilities measured by the subtests.

- There were small changes in group variability between pre and post-testing in both subtests. Children in school CES 235 finished first grade being a heterogeneous group according to their scores in English Vocabulary and Comprehension and they finished second grade with a similar degree of heterogeneity.

- The correlation between raw scores in Spanish Vocabulary and Comprehension was higher. A coefficient of .94 means that about 88 percent of the scores in one of the subtests can be predicted from the scores in the other subtest. Changes in one of the subtests will be followed by similar changes, to a high degree, in the other subtest. The coefficient was greater than the coefficients reported in the manual which are in the 80's.

c. Tests of Basic Language Competence

Tests of Language Competence in Spanish and English were given to all children who were judged by the teachers to be minimally competent in Spanish or English. Only subtests 1 and 2 were given. Different forms were used in Spanish and in English.

Since these tests have not been normed on a national sample, percentiles ranks are not included here. The only possible comparison is to a perfect total score which is reported at the bottom of each table.

The test of Language Competence was designed to measure a child's ability to understand Spanish and English. Children who speak one or both languages as a native are expected to make near perfect scores on the subtests.

The subtests, Oral Vocabulary, presents words to the child from which the child must identify the corresponding picture.

Table 22 has the pertinent information about the results of this test. From data in this table we can conclude that:

Table 22: Oral Vocabulary: Means and Standard Deviations on Spanish and English Versions of the Language Competence Test for the Second Grade.

Language	Pre-test[1]			Post-test			Mean Gain	t-test[2] Values
	N	Mean	SD	N	Mean	SD		
Spanish	46	17.80	8.09	51	20.05	6.20	2.25	1.56
English	46	22.15	3.25	53	21.69	4.50		.58

[1]Test scores are raw scores.

[2]Non-significant at the .05 level.

NOTE: Perfect score = 24.

- Children from school CES 235 scored relatively low both in Spanish and in English in the pre-test. Children finished first grade with a knowledge of English and Spanish words that was less than perfect.

- Children had a better knowledge of English than Spanish words at the end of the first grade. Their mean score in English was closer to a perfect score than the mean score in Spanish.

- Children formed a heterogeneous group according to their scores in Spanish as indicated by the large size of the standard deviation in the pre-test. In order to include 96 percent of the scores in Spanish a range of about 32 points would be needed which is not possible because the highest possible range is 24. This means that most of the children had a near perfect score in Spanish since the mean is close to 24, but a few of them had a very low score.

- Children formed a more homogeneous group in English than in Spanish, according to their scores in Oral Vocabulary. Nevertheless, there was some variability in English scores.

- Children did not show any significant change in Oral Vocabulary during eleven months of school in either Spanish or English. The mean gains between pre and post-test scores were not large enough to be significant at

the .05 level as indicated by the t-test values. Both mean gains can be explained as a result of chance alone and they do not represent real changes in Vocabulary.

- During the school period children became a more heterogeneous group in English, but they became more homogeneous in Spanish. There was an increase in the standard deviation of English scores and a reduction in Spanish scores.

5. Third Grade Achievement

Third grade children were given the Metropolitan Achievement Test, Reading, the Inter-American Series Test of Reading and the Basic Language Competence test. Children were pre-tested at the end of the second grade and post-tested at the end of the third grade, using the same test level in both situations.

a. MAT-Reading

Children were tested in Word Knowledge and Reading subtests from the MAT, Primary II Level.

The 40 items of the Word Knowledge subtest measure the extent of pupils' reading vocabulary. Seventeen items are in the word-picture association format. Twenty-three items require pupils to identify a synonym, antonym, or classification for a given word.

The 44 items of the Reading subtest measure pupils' comprehension of written material. Thirteen items require pupils to select one of three sentences which best describes a given picture. Thirty-one items require pupils to read a paragraph and answer questions about what they have read.

1) Word Knowledge

The information related to the Word Knowledge subtest is reported in Table 23. This table contains information about the number of students and mean grade equivalent on the pre-test, post-test, gain between pre and post-test, and difference between predicted and actual post-test.

Table 23. Word Knowledge: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Third Grade.

	Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2] Mean Gain	t-test Values[2] Difference
N	40	40					
Mean	2.30	2.85	.55	3.09	-.24	2.97	1.03
Standard Deviation	.71	.93		1.14			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 2.80 at the time of the pre-test and 3.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children, as a group, scored below the national norms when they were pre-tested. They scored lower than expected of them if they comformed to the national sample. The result, though, was expected given the special characteristics of school CES 235.

- The mean grade equivalent was 5 months below the national level. This means that children in the program finished second grade with achievement scores lower than the national norms.

- There were differences among the students in their Word Knowledge scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample, range of 1 year, 4 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

The column labeled "Mean Gain" merely indicates the difference between the pre-test and the post-test mean scores. From this column we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 5 and 1/2 months higher than their mean score a year earlier. The gain was significant beyond the .05 level as indicated by the t-test value, and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 5 and 1/2 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 5 and 1/2 months lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national

sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished one year of school being a heterogeneous group, and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

The column labeled "Predicted Post-test" indicates a statistical calculation based on pre-test scores. Each student's achievement thus far in school is divided by the number of months he has been in school (excepting kindergarten) and from this monthly mean gain the predicted post-test scores can be determined by simply multiplying the monthly mean gain by the number of months in the school between pre and post-testing and adding the result to the pre-test scores. The reason for this statistical manipulation is to determine whether the children gain more during the period in the program than they would have if they had continued making progress at the same rate they have thus far in their school career.

It should be noted that the predicted post-test scores are a direct reflection of the pre-test scores. If pre-test scores are low, a sample has gained very little up to that point and would be expected to gain little in the future without some special treatment. None of the school samples were expected to gain to their grade level if they were not at that level when they were pre-tested.

The column labeled "Difference" is the most important of the chart. If indicates the difference between what the children were expected to achieve and what they did achieve. From that column, the following conclusion can be drawn:

- The difference between actual and predicted post-test scores was negative. Children's mean score at the end of the year was lower than the mean expectations based on their scores at the end of the previous year. The school's impact was not beneficial for the students as a group. During the year, students improved at a rate which was smaller than

the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test was not significant at the .05 level as indicated by the t-test value. The negative change of about 2 months below expectations that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the loss was a real loss in the ability measured by the test and it can be explained by chance alone.

2) Reading

All the information about the Reading subtest is reported in Table 24. The information is reported in a similar form to the form used in the previous table.

Table 24. Reading: Means and Standard Deviation for the Pre-test Actual Post-test, Predicted Post-test, Mean Gain between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Third Grade.

	1 Pre-test	Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values 2 Mean Gain	Difference
N	40	40					
Mean	2.10	2.62	.52	2.77	-.15	3.03*	.71
Standard Deviation	.61	.91		.98			

[1]The test scores are grade equivalents.

[2]*Significant beyond the 105 level.

Note: The National Grade Level was 2.80 at the time of the pre-test and 3.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, again, was expected given the special characteristics of school

- The mean grade equivalent was 7 months below the national level. This means that children in the program finished second grade with achievement scores lower than the national norms.

- There were differences among the students in their Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 1 year, 2 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "mean gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored about 5 months higher than their mean score a year earlier. The gain was significant beyond the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 6 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 6 months lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished one year of school being a heterogeneous group, and they finished with a greater degree of heterogenity

than they had a year before.

c) Program Impact

From the column labeled "Difference" we may conclude that:

- The difference between actual and predicted post-test scores was negative. Children's mean score at the end of the year was lower than the mean expectation based on their scores at the end of the previous year. The school's impact was not beneficial for the students as a group; during the year, students improved at a rate which was smaller than the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test was not significant at the .05 level as indicated by the t-test value. The negative change of about 1 and 1/5 months below expectations that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the loss was a real loss in the ability measured by the test and it can be explained by chance alone.

3) Total Reading

The information related to the Total Reading subtest is reported in Table 25. The information is reported in a form similar to the form used in previous tables.

Table 25: Total Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Third Grade.

	Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2] Mean Gain	t-test Values[2] Difference
N	40	40					
Mean	2.20	2.75	.55	2.93	-.18	3.09*	.97
Standard Deviation	.62	.94		1.00			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 2.80 at the time of the pre-test and 3.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, again, was expected given the special characteristics of school CES 235.
- The mean grade equivalent was 5 months below the national level. This means that children in the program finished second grade with achievement scores slightly lower than the national norms.
- There were differences among the students in their Total Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 1 year, 2 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "Mean Gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 5 and 1/2 months higher than their mean score a year earlier. The mean gain was significant beyond the .05 level as indicated by the t-test value; it could not be explained by chance alone; it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 5 and 1/2 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 5 and 1/5 months lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished one year of school being a heterogenous group and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

From the column labeled "Difference" the following conclusions are clear:

- The difference between actual and predicted post-test scores was negative. Children's mean score at the end of the year was lower than the mean expectation based on their scores at the end of the previous year. The school's impact was not beneficial for the students as a group; during the year, students improved at a rate which was smaller than the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test, however, was not significant at the .05 level as indicated by the t-test value. The loss of about 2 months below expectations that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the loss was a real loss in the ability measured by the test and it can be explained by chance alone.

4) Comparison of Subtests

Table 26 includes information about each subtest and the total test. Information is arranged in a form similar to the form used in the previous tables. From this table we have the following additional information:

Table 26: Means and Standard Deviation for the Pre-test, Actual Post-test, Predicted Post-test and t-test Values for the Difference Between Pre-test, Actual Post-test and Predicted Post-test on each Subtest of the MAT-Reading for the Third Grade.

Test[4]	Pre-test[1]			Actual Post-test			Mean Gain	Predicted Post-test		Difference	t-test Values	
	N	Mean	SD	N	Mean	SD		Mean	SD		2	3
Word Knowledge	40	2.30	.71	40	2.85	.93	.55	3.09	1.14	-.24	2.97*	1.03
Reading	40	2.10	.61	40	2.62	.91	.52	2.77	.98	-.15	3.03*	.71
Total Reading	40	2.20	.62	40	2.75	.94	.55	2.93	1.00	-.18	3.09*	.97

[1]The test scores are grade equivalent.

[2]t-test values for the difference between pre and actual post-test mean-scores.

[3]t-test values for the difference between actual and predicted post-test mean scores.

*t-test values significant beyond the .05 level.

[4]Correlation between subtest: r= .62.

NOTE: The National Grade Level was 2.80 at the time of the pre-test and 3.90 at the time of the post-test.

a) Pre-program Level

From pre-test scores it can be said that:

- Children scored below national norms in both subtests, Word Knowledge and Reading.
- Children finished the second grade 5 months below national norms in Word Knowledge and 7 months below national norms in Reading.
- The gap between grade equivalent and national norms was similar in both subtest; students scored 2 months lower in Reading than in Word Knowledge.
- Students were heterogeneous groups in both subtest according to their scores. The degree of heterogeneity was similar in both subtests.

b) Pre-post Test Change

From the column labeled "Mean gain" the following conclusions can be drawn:

- There was a positive gain in grade equivalent between pre and post-test. Children, after eleven months in school, scored nigher than when they finished the year before. They improved in both subtest.
- Children improved about the same amount in Word Knowledge and Reading, the difference being less than 1/2 month.
- The gap between mean grade equivalent and national norms increased in Word Knowledge and in Reading. During eleven months of school, children in the sample improved 5 months in Reading and 5 and 1/2 months in Word Knowledge, thus increasing the gap by 5 and 1/2 months in Word Knowledge and 6 months in Reading.
- There were substantial increases in group heterogeneity in both subtests as indicated by the size of the standard deviations for the pre and post-test.

c) Program Impact

From column labeled "Difference" it can be concluded that:

- There was a negative difference in Word Knowledge and Reading between predicted and actual post-test,.
- The program's impact was not beneficial in Word Knowledge, in Reading, or in Total Reading. Students scored lower than was expected of them with expectations based on their part improvement.
- Change beyond expectations was not significant in either subtest. The changes in Word Knowledge, Reading and Total Reading were below expectations and can be explained as a result of chance alone.
- The correlation between subtests was moderate with a coefficient of .62, about 36 percent of the raw scores in one of the subtest can be predicted from the scores on the other subtest. Changes in one subtest are followed by substantial and similar changes in the other subtest.

b) Inter-American Series Test of Reading

The test of Reading in English was given to all children, either Spanish or English dominant. The test of Reading in Spanish was given only to those students with some ability in Spanish. Level II was used with third grade children. Form CE was used in English and form DEs in Spanish for the post-test. Form CE in English was used in the pre-test, the only language in which the second grade children were tested.

Caution should be used against attributing too much significance to percentile ranks as this test was normed on small, specific groups of children. These groups were, in addition, only tested in their dominant language whereas the sample at C.E.S. 235 was tested in their secondary language as well.

The Vocabulary subtest measures the child's ability to match pictures with words. The Comprehension subtest measures the student's ability to relate sentences and stories to pictures. The total is the sum of the three subtests.

Comprehension at Level II is measured by two subtests, Speed of Comprehension and Level of Comprehension. Both subtests have the same format; the only difference is the time allowed to answer each one of them. A child with mastery of the language can answer correctly all the items on the Level of Comprehension during the time allowed; in Speed of Comprehension, on the other hand, no one can answer all the items because there is not time enough to read them.

1) Test of Reading in Spanish

Table 27 has the information about the means, standard deviations and percentile ranks for the post-test scores. Children were not pre-tested and, therefore, no information about progress during one year in the program is available. From this table we have the following important information:

Table 27. Means and Standard Deviations for the Post-test, and Percentile Ranks of each Subtest of the Inter-American Series Test of Reading, in Spanish, Level II, for the Third Grade.

Test	Post-test[1]			[2] Percentiles
	N	Mean	SD	Post-test
Vocabulary	8	21.50	12.42	37
Speed of Comprehension	8	9.25	7.68	51
Level of Comprehension	8	18.75	7.61	42
Total Reading	8	49.50	25.04	39

[1] Test scores are raw scores.

[2] Percentiles are for third grade in Puerto Rico during the fall semester.

a) Analysis of Subtest Scores

- Children in school C.E.S. 235 scored below the norms for third grade children, fall semester, in Puerto Rico in Vocabulary, Level of Comprehension and Total Reading; they scored the same as the norms in Speed of Comprehension. Children in our sample had less knowledge of Spanish words and a lower level of Comprehension of statements

written in Spanish than children of the same school level in Puerto Rico; there was no difference between both samples in Speed of Comprehension of statements written in Spanish. It must be remarked that the norms are for the fall semester while children in our sample were tested at the end of the spring semester; third grade children in school C.E.S. 235 were older than the normed sample from Puerto Rico.

- Children in our sample were expected to score lower than the normed sample and they did in Vocabulary and Level of Comprehension. Cultural and language differences between Puerto Rican children living on the island and Spanish speaking children living in New York can explain the differences in Spanish scores.

- It was not expected that children in our sample would score as high or higher than the normed sample. The finding is remarkable. A test of pure speed is supposed to be very easy and each item must be answered correctly by every child if he has enough time; Speed of Comprehension measures nothing else but speed. The difference between Puerto Rican children and Spanish children from New York is not in Speed of Comprehension of easy statements but in Vocabulary and Level of Comprehension; our sample had a lack of Spanish vocabulary and also a lack of ability to understand written statements if they were not very easy.

- There were a great variety of scores in Vocabulary, Speed and Level of Comprehension as indicated by the size of the standard deviations which can be considered as large. In order to include 96 percent of the scores a range of more than 30 points was needed in Speed and Level of Comprehension and more than 40 points in Vocabulary; the greatest possible range is 40 in Vocabulary level and 30 in Speed of Comprehension. There were large differences among students in the sample in their mastery of Spanish.

b) Correlations Between The Subtest

Table 28 reports the coefficients of correlation between the three subtests for school C.E.S. 235 children and for the normed sample. From this table we can conclude that:

Table 28. Product-Moment Correlation Coefficients Between Vocabulary, Speed and Level of Comprehension on the American Series Test of Reading, in Spanish, Level II, for the Third Grade.

	C.E.S. 235	National Sample
Vocabulary vs. Speed of Comprehension	.57	.34
Vocabulary vs. Level of Comprehension	.80	.60
Speed of Comprehension vs. Level of Comprehension	.80	.62

- There was a moderate correlation between Vocabulary and Speed of Comprehension. About 32 percent of scores on one of the subtests can be explained by the scores on the other subtest. About a third of the scores in Speed and Comprehension was the result of vocabulary knowledge.

- There was a high correlation between Vocabulary and Level of Comprehension and between Speed and Level of Comprehension about 64 percent of the scores in Level of Comprehension can be accounted for by knowledge of words and also about 64 percent of the scores in Speed of Comprehension are related to scores in Level of Comprehension.

- All the correlation coefficients are higher in our sample than in the national norms. The interrelation among the subtests was greater in our sample than in the normed samples. It is also true that progress in one of the subtests was followed by greater progress in the other subtests in our sample than in the normed sample.

- Vocabulary is important for comprehension of written statement, both in Speed and Level of Comprehension, but it is more important for Level of Comprehension than for Speed of Comprehension. The finding is understandable because, as we said earlier, a speed test must have very easy items and, therefore, the differences in speed scores are not the result of knowledge of words but the consequence of other language ability. In Level of Comprehension, on the other hand, the scores depend heavily on vocabulary knowledge.

- A recommendation based on the information in this table is that it would be useful to teach vocabulary so that children may be able to understand Spanish better. But the correlative between subtests is not perfect and, therefore, children must also be taught the technique of standard written statements beyond the knowledge of each word.

2) Test of Reading in English

Careful note must be made that the post-test percentiles discussed in the following tables are related to the mean raw scores for the post-test. The post-test, as stated earlier, was given to the children during spring semester of their third grade. With some children scoring well below the national norms, we confront two problems: 1) did they improve at all during the program year, and 2) if they made progress, did they improve as much as the national sample. In order to answer the first question, the mean raw score of the post-test was converted into percentile ranking at the same level as the pre-test (i.e., second grade). To answer the second question, the mean raw score must be compared with that of the third grade national sample.

a) Vocabulary

Table 29 contains the information about the vocabulary subtest of the Test of Reading in English. The information is reported in a form similar to the form used in the Spanish version of the same test. The following conclusions are drawn from this table.

Table 29: Vocabulary: Means and Standard Deviations for the Pre-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level II, for the Third Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles[2] Pre-test 2nd Grade	Post-test 2nd Grade	Post-test 3rd Grade	t-test[3] Values
N	41	38					
Mean	19.30	25.44	6.14	50	76	33	3.17
Standard Deviation	8.50	8.71					

[1]Test scores are raw scores.

[2]Percentiles are for the Second and Third Grades during the fall semester.

[3]t-test values = 3.17 which was significant beyond the .05 level.

- Children in school C.E.S. 235 finished second grade with a mean score in English Vocabulary that was at exactly the 50th percentile for native English speaking second grade children as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of 50 percent of second grade students from a national sample. This result was not expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the English vocabulary score. The result though, can be explained by the fact that the national sample was normed during the fall and our sample was tested during the spring.

- There was a great variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 34 points would be needed in order to include 96 percent of the scores in the sample; 40 is the possible range of scores.

- During eleven months of school, children improved in their English Vocabulary scores. The mean score increased by 26 percentile points. It must be added that the norms are for second grade children, fall semester, and our sample was post-tested during the spring semester.

- Children in C.E.S. 235 did not improve as much as the national sample. Children from our sample finished second grade at the same level as a national second grade sample; but at the end of the third grade they were 17 percentile points below the national third grade level, thus showing an improvement substantially smaller than what is expected from a national sample.

- The mean gain of 37 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in English Vocabulary was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a small increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores in English Vocabulary at the end of second grade and they had a similar degree of heterogeneity at the end of the third grade.

b) Speed of Comprehension

Table 30 contains the information about the Speed of Comprehension subtest of the Test of Reading in English. The information is reported in a form similar to the form used in the previous table. The following conclusions are drawn from this table:

Table 30. Speed of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level II for the Third Grade.

	1 Pre-test	Post-test	Mean gain	Percentiles 2 Pre-test 2nd grade	Post-test 2nd grade	Post-test 3rd grade	3 t-test values
N	38						
Mean	7.70	12.89	5.19	70	95	48	3.78
Standard Deviation	4.50	7.18					

[1] Test scores are raw scores.

[2] Percentiles are for the second and third grades during the fall semester.

[3] t-test value=3.78 which was significant beyond the .05 level.

- Children in school C.E.S. 235 finished second grade with a mean score in English Speed of Comprehension that was above the 50th percentile for native English speaking second grade children as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of 70 percent of second grade students from a national sample. The result, again, was not expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in English Comprehension score. The result, though, can be explained by the fact that the national sample was normed during the fall and our sample was tested during the spring.

- There were a great variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 18 points would be needed in order to include 96 percent of the scores in the sample; 30 is the possible range of scores.

- During eleven months of school, children improved in their English Speed of Comprehension scores. The mean score increased by 25 percentile points. It must be added that the norms are for second grade children, fall semester, and our sample was post-tested during the spring semester.

- Children in school C.E.S. 235 did not improve as much as the national sample. Children from our sample finished second grade at a higher level than the level for the national second grade sample, but at the end of the third grade they were 2 percentile points below the national third grade level, thus dropping a substantial number of percentile points from the national level at which they were a year earlier.

- The mean gain of 25 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in English Speed of Comprehension was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a substantial increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores in English Speed of Comprehension at the end of second grade and they had a greater degree of heterogeneity at the end of the third grade.

c) Level of Comprehension

Table 31 contains the information about the Level of Comprehension subtest of the Test of Reading in English. The information is reported in a form similar to the form used in the previous tables. The following conclusions are drawn from this table:

Table 31. Level of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level II for the Third Grade.

	1 Pre-test	Post-test	Mean gain	Percentiles 2 Pre-test 2nd grade	Post-test 2nd grade	Post-test 3rd grade	3 t-test values
N	41	38					
Mean	16.00	24.89	8.89	50	90	40	5.47
Standard Deviation	6.30	8.10					

[1]Test scores are raw scores.

[2]Percentiles are for the Second and Third grades during the fall semester.

[3]t-test value=5.47 which was significant beyond the .05 level.

- Children in school C.E.S. 235 finished second grade with a mean score in English Level of Comprehension that was at exactly the 50th pecentile for native English speaking second grade children as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of 50 percent of second grade students from the national sample. The result was not expected: cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in English Level of Comprehension scores. The result, however, can be explained by the fact that the national sample was normed during the fall and our sample was tested during the spring.

- There were a great variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 25 points would be needed in order to include 96 percent of the scores in the sample; 40 is the possible range of scores.

- During eleven months of school, children improved in their English Level of Comprehension scores. The mean score increased by 40 percentile points. It must be added that the norms are for second grade children, fall semester, and our sample was post-tested during the spring semester.

- Children in school C.E.S. 235 did not improve as much as the national sample. Children from our sample finished second grade at the same level as the national second grade sample, but at the end of the third grade they were 10 percentile points below the national third grade level, thus showing an improvement substantially smaller than what is expected from a national sample.

- The mean gain of 40 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in English Level of Comprehension was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a substanial increase in group variability between pre and post-testing as indicated by the size of the standard deviation. Children were a heterogeneous group according to their scores in English Level of Comprehension at the end of second grade and they had a greater degree of heterogeneity at the end of the third grade.

d) Total Reading

Table 32 contains the information related to English Total Reading in a form that is similar to the form used in previous tables. From data in this table we can draw the following conclusions:

Table 32. Total Reading: Means and Standard Deviations for the Pre-test and Percentile Ranks on the Inter-American Series Test of Reading, in English, Level II for the Third Grade.

	[1] Pre-test	Post-test	Mean gain	Percentiles [2] Pre-test 2nd grade	Post-test 2nd grade	Post-test 3rd grade	[3] t-test values
N	38	38					
Mean	42.40	63.23	21.83	55	87	37	5.11
Standard Deviation	15.80	21.09					

[1] Test scores are raw scores.

[2] Percentiles are for second or third grades during the fall semester.

[3] t-test value=5.11 which was significant beyond the .05 level.

e) Comparison of Subtests

Table 33 includes the information related to each subtest and the total test. From this table we draw the following conclusions:

Table 33. Means and Standard Deviations for the Pre-test, Post-test, Percentile Ranks of each Subtest of the Inter-American Series Test of Reading, in English, Level II, for the Third Grade.

Test	Pre-test[1]			Post-test			Mean Gain	Percentiles[2]			t-test Values[3]
								Pre-test	Post-test		
	N	Mean	SD	N	Mean	SD		2nd	2nd	3rd	
Vocabulary	41	19.30	8.50	38	25.44	8.71	6.14	50	76	33	3.17*
Speed of Comprehension	38	7.70	4.50	38	12.89	7.18	5.19	70	95	48	3.78*
Levle of Comprehension	41	16.00	6.30	38	24.89	8.10	8.89	50	90	40	5.47*
Total Reading	38	42.40	15.80	38	36.23	21.09	21.83	55	87	37	5.11*

[1]Test scores are raw scores.

[2]Percentiles are for second or third grades during the fall semester

[3]*Significant beyond the .05 level.

- Children in school C.E.S. 235 finished second grade, as reported by the pre-test data, with a mean score in English Vocabulary, Speed and Level of Comprehension, and Total Reading as high or higher than the 50th percentile of the native English speaking second grade children. It must be emphasized again that the normed sample was tested during the fall and children form school C.E.S. 235 during the spring semester.

- There was a **range of scores in every subtest as** indicated by the size of the standard deviation. The group was a heterogeneous group and it had different degrees of heterogeneity on each subtest. The greatest range of scores occurred in Vocabulary.

- During eleven months of school children improved in every subtest. The greatest improvement was in Level of Comprehension where the difference between pre and post-test mean scores was 40 percentile points.

- Children in school CES 235 finished second grade with a mean score in English Total Reading that was above the 50th percentile for native English speaking second grade children. Only 46 percent of second grade children from a national sample scored higher than the mean score for the children in our sample. The result was not expected, but it can be explained as a result of the different time in which the normed sample and the children in school CES 235 were tested; the national sample was tested during the fall and our sample, during the spring semester.

- There was a great variety of scores in English Total Reading within the group as indicated by the size of the standard deviation for the pre-test. A range of 63 points would be needed to include 96 percent of the scores in the sample; the possible range is 110 points.

- During eleven months of school, children improved in their English Reading scores. The mean score increased by 32 percentile points. It must be added that the norms are for second grade children, fall semester, and our sample was post-tested during the spring semester.

- Children in school C.E.S. 235 did not improve as much as a national sample. Children from our sample finished second grade a little higher than the level of a national second grade sample; but at the end of the third grade they were 13 percentile points below the national third grade level thus showing an improvement substantially smaller than what is expected from a national sample.

- The mean gain of 32 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in English Total Reading was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a substantial increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores in English Reading at the end of second grade and they had a greater degree of heterogeneity at the end of the third grade.

- The mean gains in every subtest and Total Reading were significant beyond the .05 level as indicated by the t-test values. The improvement were greater than what could be expected by chance alone and they represented real improvements in the abilities measured by the subtests.

- Children in school C.E.S. 235 did not improve in any of the subtests as much as the national sample. Children from our sample scored as high or higher than the 50th percentile for a national sample at the end of the second grade, but a year later they scored below the 50th percentile of the third grade in every subtest even though the national sample was tested during the fall and our sample during the spring semester.

f) Correlation Between Subtests

Table 34 reports the correlations between the three subtests for school C.E.S. 235 and for the national sample.

Table 34: Product-Moment Correlation Coefficients Between Vocabulary, Speed and Level of Comprehension on the Inter-American Series Test of Reading, in English, Level II, for the Third Grade.

	C.E.S. 235	National Sample
Vocabulary vs. Speed of Comprehension	.59	.55
Vocabulary vs. Level of Comprehension	.70	.58
Speed of Comprehension vs. Level of Comprehension	.67	.74

From this table we have the following conclusions:

- The correlations between subtests were moderate in school C.E.S. 235. This means that changes in scores on one of the subtests were followed by similar changes in other subtests. The amount of change in a subtest that could be explained by changes in another subtest were as follows: about 35 percent be- Vocabulary and Level of Comprehension, and 45 percent between Speed of Comprehension and Level of Comprehension.

- The correlation coefficients between Vocabulary and Speed and Speed and Level of Comprehension were a little higher in school C.E.S. 235 than in the national sample. This means that improvement in English Vocabulary was more related in our sample than in a national sample to improvements in Speed and Level of Comprehension. The opposite is also true; low scores in Speed and Level of Comprehension were the result of low scores in Vocabulary. This connection was stronger in our sample than in a national sample.

- The correlation between Speed and Level of Comprehension was a little lower in school C.E.S. 235 than in a national sample. The two reading skills were relatively more independent in our sample than in a national sample.

c. Test of Basic Language Competence

Tests of Language Competence in Spanish and English were given to all children who were judged by the teachers to be minimally competent in Spanish and English. Only subtests 1 and 2 were given. Different forms were used in Spanish and in English.

Since these tests have not been normed on a national sample, percentile ranks are not included here. The only possible comparison is to a perfect total score which is reported at the bottom of each table.

The test of Language Competence was designed to measure a child's ability to understand Spanish and English. Children who speak one or both languages as natives are expected to make near perfect scores on the subtests.

The subtest 1, Oral Vocabulary, presents words to the child from which the child must identify the corresponding picture. The subtest 2, Phonemic Discrimination, measures phonemic discrimination at word and phrase level; the test requires the child to discriminate whether two words sound the same or different.

1) Oral Vocabulary

Table 35 has the pertinent information about the Oral Vocabulary in Spanish and in English. From the data on this table we may conclude that:

Table 35. Oral Vocabulary: Mean and Standard Deviations on the Spanish and English Versions of the Language Competence Test for the Third Grade.

Language	Pre-test [1]			Post-test			Mean Gain	t-test [2] Values
	N	Mean	SD	N	Mean	SD		
Spanish	32	16.40	6.80	23	16.65	8.44	.25	.15
English	48	23.80	.40	22	23.72	.55	-.08	.69

[1] Test scores are raw scores.

[2] Non-significant beyond the .05 level.

NOTE: Perfect score = 24.

- Children from school C.E.S.235 scored ver low in Spanish in the pre-test. Children finished second grade with a knowledge of Spanish words that was less than perfect.

- Students formed a heterogeneous group according to their scores in Spanish in the pre-test. The highest possible range is 24 but a range of more than 27 is needed in order to include only 96 percent of the scores in Spanish. This means that the sample was made up of two different groups, one with high scores and the other with very low scores with few student scoring in between.

- Children had practically a perfect mean score in English. They finished second grade with a perfect knowledge of simple English words.

- Children finished second grade being a homogeneous group according to their scores in English. Very few students scored below a perfect score and those who did were not far below.

- Children did not change significantly in English Oral Vocabulary during eleven months of school. There was not too much room for change since must of them finished second grade with a perfect score. The negative mean change of .08 between pre and post-test was not significant and it is the result of chance alone. The students did not change substantially in the group's homogeneity either.

- Students did not show any improvement in Spanish during the eleven months of school. The positive mean gain of .25 was not large enough to be beyond changes expected by chance alone.

- Children because a more heterogeneous group in their Spanish scores during the school year as indicated by a substantial increase in the size of the standard deviation between the pre and post-testing.

2) Phonemic Discrimination

Table 36 reports the pertinent information about the subtest phonemic discrimination in Spanish and in English. From this table we have the following information.

Table 36. Phonemic Discrimination: Means and Standard Deviations on Spanish and English Version of the Language Competence Test for the Third Grade.

Language	Pre-test [1]			Post-test			Mean Gain	[2] t-test Values
	N	Mean	SD	N	Mean	SD		
Spanish	22	25.90	7.90	21	18.52	3.44	-7.38	3.94*
English	24	33.90	6.00	23	31.95	7.38	-1.95	1.00

[1] Test scores are raw scores.

[2] *Significant beyond the .05 level.

Note: Perfect score = 40.

- Children from school C.E.S. 235 scored low in Spanish and in English in the pre-test. Their mean score at the end of second grade was less than perfect. Their mastery of Spanish and English sounds was less than that expected from children that speak either language as a native language. Students from our sample had a language disadvantage in relation to both Spanish and English speaking students.

- Children had a better discrimination of English than Spanish sounds as shown by the difference in mean score for the pre-test.

- Students were a heterogeneous group according to their scores in Spanish and in English as indicated by the size of the standard deviation for the pre-test. There was a wide range of scores in both languages.

- During eleven months of school, children decreased their mean score in Spanish and in English as shown by the negative mean gain. The negative change, though, was significat beyond the .05 level only in Spanish. The change they showed in English can be a chance effect, but in Spanish it was a real negative change indicating a deterioration of their ability to discriminate sounds.

- There was a substantial decrease in group heterogeneity in Spanish scores; students, as a group, snarply reduced the differences among them. In English there was an increase in group heterogeneity; the range of scores and the differences between students increased during the year.

6. Fourth Grade Achievement

Fourth grade children were given the Metropolitan Achievement Test, Reading, the Inter-American Series Test of Reading and the Basic Language Competence Test. Children were pre-tested at the end of the third grade and post-tested at the end of the fourth grade, using if possible, the same test level in both situations.

a. MAT - Reading

Children were tested in Word Knowledge and Reading subtests from the MAT, Elementary level.

The 50 items of the Word Knowledge subtest measures the extent of pupils' reading vocabulary. Items require pupils to identify synonyms, antonyms, or word classification. Items range from primary level to junior high level in difficulty.

The 45 items of the Reading subtest measure pupils' ability to comprehend written material. Pupils read a paragraph and then answer questions about it. Items cover comprehending literal meanings of passages, drawing inferences from the material, identifying main ideas, and determining word meanings from context.

1) Word Knowledge

The information related to the Word Knowledge subtest is reported in Table 37. This table contains information about the number of students and mean grade equivalent on the pre-test, post-test, gain between pre and post-test and difference between predicted and actual post-test.

Table 37: Word Knowledge: Means and Standard Deviations for the Pre-test, Post-test, Predicted Post-test, Mean Gain between Pre and Post-test, Difference between Predicted and Actual Post-test and t-test Values on the MAT for the Fourth Grade.

	Pre-test[1]	Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2]	
						Gain	Difference
N	33	44					
Mean	2.40	2.91	.51	2.95	-.04	2.60*	.17
Standard Deviation	.98	.74		1.36			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 3.80 at the time of the pre-test and 4.90 at the time of the post-test

a) Pre-program Level

The data for pre-test scores as indicated in the table show the following:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, again, was expected given the special characteristics of school CES 235.

- The mean grade equivalent was 1 year, 5 months below the national level. This means that children in the program finished third grade with achievement scores lower than the national norms for second grade.

- There were differences among the students in their Total Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 1 year, 7 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "Mean Gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 5 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 6 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school, the students were 6 months lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial decrease in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished the year of school being a heterogeneous group, but they finished with a smaller degree of heterogeneity than they had a year before.

c) Program Impact

The column labeled "Predicted Post-test" indicates a statistical calculation based on pre-test scores. Each student's achievement thus far in school is divided by the number of months he has been in school (excepting kindergarten) and from this monthly mean gain the predicted post-test scores can be determined by simply multiplying the monthly mean gain by the number of months in the school between pre and post-testing and adding the result to the pre-test scores. The reason for this statistical manipulation is to determine whether the children gain more during the period in the program than they would have if they had continued making progress at the same rate they have thus far in their school career.

It should be noted that the predicted post-test scores are a direct reflection of the pre-test scores. If pre-test scores are low, a sample has gained very little up to that point and would be expected to gain little in the future without some special treatment. None of the school samples were expected to gain to their grade level if they were not at that level when they were pre-tested.

The column labeled "Difference" is the most important of the chart. It indicates the difference between what the children were expected to achieve and what they did achieve. From that column, the following conclusions can be drawn.

- The difference between actual and predicted post-test scores was negative. Children's mean scores at the end of the year was lower than the mean expectations based on their scores at the end of the previous year. The school's impact was not beneficial beyond expectations for the students as a group; during the year, students improved at a rate which was slightly smaller than the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test was not significant at the .05 level as indicated by the t-test value. The negative difference of about 1/2 month beyond expectations that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the loss was a real loss in the ability measured by the test and it can be explained by chance alone.

2) Reading

The information about the reading subtest is reported in Table 38. The information is reported on a form similar to the form used in the previous table.

Table 38: Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain between Pre and Post-test, Difference between Predicted and Actual Post-test and t-test Values on the TNT for the Fourth Grade.

	Pre-test[1]	Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2] Mean	Difference
N	33	44					
Mean	2.20	2.85	.65	2.67	.18	2.94*	.65
Standard Deviation	.84	1.04		1.17			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 3.80 at the time of the pre-test and 4.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, again, was expected given the special characteristics of school CES 235.

- The mean grade equivalent was 1 year 6 months below the national level. This means that children in the program finished third grade with achievement scores lower than the national norms for second grade.

- There were differences among the students in their reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 1 year, 6 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "Mean Gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 6 and 1/2 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 4 and 1/2 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school, the students were 4 and 1/2 months lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished the year of school being a heterogeneous group and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

From the column labeled "Difference" we draw the following conclusions:

- The difference between actual and predicted post-test scores was positive. The children's mean score at the end of the year was higher than the mean expectation based on their scores at the end of the previous year. The school's impact was beneficial for the students as a group; during the year, students improved at a rate which was greater than the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test was not significant at the .05 level as indicated by the t-test value. The positive improvement of about 2 months beyond expectations that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the improvement was a real improvement in the ability measured by the test and it can be explained by chance alone.

3) Total Reading

Table 39 has the information about Total Reading. The information is reported in a form similar to the form used in previous tables.

Table 39: Total Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain between Pre and Post-test, Difference between Predicted and Actual Post-test t-test Values on the MAT for the Fourth Grade.

	Pre-test[1]	Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values	
						Mean Gain	Difference
N	33	44					
Mean	2.30	2.83	.53	2.81	.02	2.72*	.09
Standard Deviation	.89	.81		1.24			

[1]The test scores are grade equivalents

[2]* Significant beyond the .05 level.

NOTE: The National Grade Level was 3.80 at the time of the pre-test and 4.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored well below the national norms when they were pre-tested. They scored lower than expected of them if conformed to the national sample. The result, though, was expected given the special characteristics of school CES 235.

- The mean grade equivalent was 1 year, 4 months below the national level. This means that children in the program finished third grade with achievement scores lower than the national norms for second grade.

- There were differences among the students in their word knowledge scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 1 year, 9 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Change

The column labeled "Mean Gain" merely indicates the difference between the pre-test and the post-test mean scores. From this column we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 5 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 6 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 6 months lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was no change in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished the year of school being a heterogeneous group and they finished with a similar degree of heterogeneity to the one they had the previous year.

c) Program Impact

From the column labeled "Difference" we conclude that:

- The difference between actual and predicted post-test scores was practically zero. The children's mean score at the end of the year was not different from the mean expectation based on their scores at the end of the previous year. The school's impact was not beneficial beyond expectations for the students as a group; during the year, students improved at a rate which was similar to the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test was not significant at the .05 level as indicated by the t-test value. The positive improvement of about .2 months beyond expectations that occurred in the students as a sample was too small to be significant beyond changes expected by chance alone.

5) Comparison of Subtests

Table 40 includes information about each subtest and the total test. Information is arranged in a form similar to the form used in the previous tables. From this table we have the following additional information.

Table 40: Means and Standard Deviation for the Pre-test, Actual Post-test, Predicted Post-test and t-test Values for the Difference Between Pre-test, Actual Post-test and Predicted Post-test on each Subtest of the MAT-Reading for the Fourth Grade.

Test[4]	Pre-test[1]			Actual Post-test			Mean Gain	Predicted Post-test		Difference	t-test Values	
	N	Mean	SD	N	Mean	SD		Mean	SD		2	3
Word Knowledge	33	2.40	.98	44	2.91	.74	.51	2.95	1.36	-.04	2.60*	.17
Reading	33	2.20	.84	44	2.85	1.04	.65	2.67	1.17	.18	2.94*	.65
Total Reading	33	2.30	.89	44	2.83	.81	.53	2.81	1.24	.02	2.72*	.09

[1]The test scores are grade equivalents.

[2]t-test values for the difference between pre and actual post-test mean scores.

[3]t-test values for the difference between actual and predicted post-test mean scores.

*t-test value significant beyond the .05 level.

[4]Correlation between subtests: r=.78.

Note: The National Grade Level was 3.80 at the time of the pre-test and 4.90 at the time of the post-test.

a) Pre-program Level

From the pre-test scores it can be said that:

- Children scored below national norms in both subtests, Word Knowledge and Reading.
- Children finished third grade one year, four months below national norms in Word Knowledge and one year, six months below national norms in Reading.
- The gap between grade equivalent and national norms was very similar in both subtests; students scored two months lower in Reading than in Word Knowledge.
- Students were heterogeneous groups in every subtest according to their scores. The degree of heterogeneity was different in both subtests; the greatest degree of heterogeneity occurred in Word Knowledge.

b) Pre-post Test Change

From the column labeled "Mean Gain" the following conclusions can be drawn:

- There was a positive gain in grade equivalent between pre and post-test. Children, after eleven months in school, scored higher than when they finished the year before. They improved in both subtests.

- Children improved less in Word Knowledge than in Reading; the difference between mean gains was one and a half months and four months.

- The gap between mean grade equivalent and national norms increased in both Word Knowledge and Reading. During eleven months of school, children in the sample improved only five months in Word Knowledge and six and a half months in Reading, thus increasing the gap by six months in Word Knowledge and by three and a half months in Reading.

- There was a substantial increase in group heterogeneity in Reading and a substantial decrease in Word Knowledge as indicated by the size of the standard deviations for the pre and post-test.

c) Program Impact

From the column labeled "Difference" it can be concluded that:

- There was a positive difference in Reading between predicted and actual post-test, but the difference was negative in Word Knowledge.

- The program's impact was beneficial in Reading and in the Total Reading. Students scored higher than was expected of them with expectations based on their past improvement.

- None of the changes beyond expectations was significant at the .05 level. All of them can be explained as a result of chance alone.

- The correlation between subtests was high. With a coefficient of .78 about 61 percent of the raw scores in one of the subtests can be predicted from the scores on the other subtest. Changes in one subtest were followed by substantial and similar changes in the other subtest.

b. Inter-American Series Test of Reading

The test of Reading in English was given to all children, either Spanish or English dominant. The test of Reading in Spanish was given only to those students with some ability in Spanish. Form CE in English and in Spanish was used for the pre-test. In the post-test form CE was used in English and form DEs in Spanish. Level II was used for the pre-test and Level III for the post-test.

Caution should be used against attributing too much significance to percentile ranks as this test was normed on small, specific groups of children. These groups were, in addition, only tested in their dominant language; whereas, the sample at C.E.S. 235 was tested in their secondary language as well. In Level II, the Vocabulary subtest measures the child's ability to match pictures with words. The Comprehension subtest measures the student's ability to relate sentences and stories to pictures. The total is the sum of the three subtests.

Comprehension at Level II is measured by two subtests, Speed of Comprehension and Level of Comprehension. Both subtests have the same format; the only difference is the time allowed to answer each one of them. A child with normal mastery of the language can answer correctly all the items on the Level of Comprehension subtest during the time allowed; in Speed of Comprehension, on the other hand, no one can answer all the items because there is not enough time to read them.

In Level III in the Vocabulary test, the item is a definition, a descriptive phrase, or a word suggesting the response; the task for the children is to find the word among different choices. In Speed of Comprehension the task is to find a word omitted from the sentence in the item; the items are arranged in cycles of similar difficulty, and the test is highly speeded. The Level of Comprehension is tested by questions based on paragraphs of generally increasing difficulty.

It is difficult to make comparisons between pre and post-test data; not only are different levels used in both situations but also the tasks are different for each level. The only possible comparison is between percentiles, but we cannot compare means or standard deviations nor can we perform the t-test.

1) Test of Reading in Spanish

a) Vocabulary

Table 41 contains the information about the means and standard deviations for the pre-test, post-test and the correspondent percentile ranks for the Spanish Vocabulary subtest.

Table 41: Vocabulary: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in Spanish, Level II and III for the Fourth Grade.

	Pre-test[1]	Post-test	Percentiles[2]		
			Pre-test	Post-test	Gain
N	36	15			
Mean	16.00	6.80	16	48	32
Standard Deviation	8.00	3.34			

[1]Test scores are raw scores.

[2]Percentiles are for the fall semester and they belong to the Third grade from Puerto Rico in the pre-test, Level II and Fourth grade from Puerto Rico in the post-test, Level III.

From this table we can draw the following information:

- Children in School C.E.S. 235 , as indicated by the pre-test data, finished the third grade with a mean score in Spanish vocabulary that was well below the norms for third grade children from Puerto Rico. About 84 percent of the third grade children in Puerto Rico scored higher than the mean score for our sample. The finding is more striking because the norms are for the fall semester and our sample children were tested during the spring semester. The result, though, was expected; cultural and language differences between Puerto Rican children living on the island and Spanish speaking children living in the United States were expected to be reflected in vocabulary scores.

- There was a great variety of scores in Spanish vocabulary at the end of the third grade as indicated by the size of the standard deviation. A range of more than 32 points is needed to include 96 percent of the sample scores. The possible range is 40. There were large differences among students in their knowledge of Spanish words as indicated by the wide range of scores.

- Students in school C.E.S. 235 finished the fourth grade slightly below the norms as indicated by the post-test scores. Only 52 percent of fourth grade children in Puerto Rico scored as high or higher than the mean score for our sample. We must remember that the norms were developed during the fall and children in school C.E.S. 235 were tested in spring, thus making our sample **at a higher level than the normed sample.**

- The sample became a more homogeneous group during the year as indicated by the relatively small standard deviation for the post-test. A range of about 13 points was needed to include 96 percent of the scores from a possible range of 45. This finding means that most of the students had a score below 13 which can be considered as a very low score.

- There was a positive improvement during eleven months of school as indicated by a gain of 32 percentile points between pre and post-test.

- A note of caution must be added. A mean score of seven or even nine can be obtained simply by guessing. In a test of 45 items with five possible answers for each item, it is probable to get a score of nine without knowing a word of Spanish and by guessing all the answers. If, in addition, a score of nine is equivalent to the 55th percentile of the normed population, it means that the test is very difficult for students; only 45 percent get a score as high or higher than the possible score they could achieve by guessing alone.

b) Speed of Comprehension

Table 42 has the information in relation to the Speed of Comprehension subtest. From this table we can conclude that:

Table 42: Speed of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in Spanish, Level II and III for the Fourth Grade.

	Pre-test[1]	Post-test	Percentiles[2]		
			Pre-test	Post-test	Gain
N	36	15			
Mean	6.40	3.00	28	41	13
Standard Deviation	2.80	1.73			

[1]Test scores are raw scores.

[2]Percentiles are for the fall semester and they belong to the:
Third grade from Puerto Rico in the pre-test, Level II.
Fourth grade from Puerto Rico in the post-test, Level III.

- Children from school C.E.S. 235, as indicated by the pre-test data, finished the third grade with a mean score in Speed of Comprehension of Spanish that was well below the norms for third grade children from Puerto Rico. About 72 percent of the third grade children from Puerto Rico scored higher than the mean score for our sample. It should be emphasized, again, that the norms were developed for the fall and our sample children were tested during the spring, thus making the difference more striking. The result was expected given the special characteristics of school C.E.S. 235.

- The sample was a homogeneous group as indicated by the small size of the standard deviation. About 96 percent of the scores are included in a range of 11 when the possible range is 30. Most of the children scored very low.

- With 40 items in the test and four choices for each item, it is quite possible to get a score of ten simply by guessing the answers; this chance score is higher than the mean score for our sample. This means that a mean score as low as the one for our sample in the pre-test is not very reliable and there is no way of knowing if it was the result of language ability or simply a guessing effect.

- The students' mean score at the end of the fourth grade was also below the norms. About 59 percent of of the fourth grade children scored higher than the mean score for our sample. The difference in time

of testing for both samples should be taken into consideration in order to interpret this result.

- The sample remained a homogeneous group after eleven months of school. Most of the students, again, did not score above the possible guessing score.

- There was a positive improvement in Speed of Comprehension of Spanish during eleven months of school as indicated by the gain of 13 percentile points between pre and post-test. But the difference is not very reliable since the mean score in both cases was below the score possible by simple guessing. It is impossible, again, to say if the gain was a gain in language ability or simply a chance effect.

c) Level of Comprehension

Table 43 contains the information about the means and standard deviations for the pre-test, post-test and the correspondent percentile ranks for the Spanish Level of Comprehension subtest.

Table 43: Level of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in Spanish, Level II and III for the Fourth Grade.

	Pre-test[1]	Post-test	Percentiles[2]		
			Pre-test	Post-test	Gain
N	36	15			
Mean	13.40	6.26	16	38	22
Standard Deviation	6.40	3.26			

[1] Test scores are raw scores.

[2] Percentiles are for fall semester and they belong to the Third grade from Puerto Rico in the pre-test, Level II and Fourth grade from Puerto Rico in the post-test, Level III.

From this table we can draw the following information:

- Children in school C.E.S. 235 as indicated by the pre-test data, finished the third grade with a mean score in Spanish Speed of Comprehension that was well below the norms for third grade children from Puerto Rico. About 84 percent of the third grade children in Puerto Rico scored higher than the mean score for

our sample. The finding, again, is striking because the norms are for the fall semester and our sample children were tested during the spring semester. The result, though, was expected, because of cultural and language differences between Puerto Rican children living on the island and Spanish speaking children living in the United States.

- There was a variety of scores in Spanish Speed of Comprehension at the end of the third grade as indicated by the size of the standard deviation. A range of more than 25 points is needed to include 96 percent of the sample scores. The possible range is 40. There were differences among students as indicated by the relatively wide range of scores.

- Students in school C.E.S. 235 finished the fourth grade below the norms as indicated by the post-test scores. About 62 percent of the fourth grade children in Puerto Rico scored as high or higher than the mean score for our sample. We must remember that the norms were developed during the fall and children in C.E.S. 235 were tested in spring, thus making our sample a higher school level than the normed sample.

- The sample became a more homogeneous group during the year as indicated by the relatively small standard deviation for the post-test. A range of about 13 points was needed to include 96 percent of the scores from a possible range of 50. This finding means that most of the students had a score below 12 which can be considered as a very low score.

- There was a positive improvement during eleven months of school as indicated by a gain of 22 percentile points between pre and post-test.

- A note of caution must be added. A mean score of seven or even ten can be obtained simply by guessing. In a test of 50 items with five possible answers for each item, it is feasible to get a score of ten without knowing a word of Spanish and by guessing all the answers. If, in addition, a score of eight is equivalent to the 50th percentile of the normed population, it means that the test is very difficult for the students; only half of the normed population reached a score higher than the possible scores they could obtain by guessing alone.

d) Total Reading

Table 44 has the information related to the Total Reading subtest. From this table we can see that:

Table 44: Total Reading: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in Spanish, Level II and III for the Fourth Grade.

	Pre-test[1]	Post-test	Percentiles[2]		
			Pre-test	Post-test	Gain
N	36	15			
Mean	35.80	16.06	15	40	25
Standard Deviation	13.90	5.58			

[1]Test scores are raw scores.

[2]Percentiles are for fall semester and they belong to the Third grade from Puerto Rico in the pre-test, Level II and Fourth grade from Puerto Rico in the post-test, Level III.

- Children in school C.E.S. 235, as indicated by the pre-test data, finished the third grade with a mean score in Spanish Total Reading that was well below the norms for third grade children from Puerto Rico. About 85 percent of the third grade children in Puerto Rico scored higher than the mean score for our sample. It must be added also that the norms are for the fall semester and our sample children were tested during the spring semester. The result, though, was expected because of cultural and language differences between Puerto Rican children living on the island and Spanish speaking children living in the United States.

- There was a great variety of scores in Spanish Vocabulary at the end of third grade as indicated by the size of the standard deviation. A range of more than 55 points is needed to include 96 percent of the sample scores. The possible range is 110. There were large differences among students in their ability to read Spanish as indicated by the wide range of scores.

- Students in school C.E.S. 235 finished the fourth grade also below the norms as indicated by the post-test scores. About 60 percent of the fourth grade children in Puerto Rico scored as high or higher than the mean score for our sample. We must remember that the norms were developed during the fall and children in school C.E.S. 235 were tested in spring, thus making our sample a higher school level than the normed sample.

- The sample became a more homogeneous group during the year as indicated by the relatively small standard deviation for the post-test. A range of about 22 was needed to include 96 percent of the scores from a possible range of 125. This finding means that most of the students had a score below 26 which can be considered as a very low score.

- There was a positive improvement during eleven months of school as indicated by a gain of 15 percentile points between pre and post-test.

- A note of caution must be indicated. The mean score is very low to be considered of much value. In a test of 125 items and five choices for each item, a score of 25 can be obtained simply by guessing. Our sample's mean score was 16, well below the possible score by guessing alone.

e) Comparison of Subtests

Table 45 has the information for each subtest of the test of Reading in Spanish. The information is reported in a form which is similar to the form used in previous tables.

Table 45: Means and Standard Deviations for the Pre-test and Post-test, Percentile Ranks of each Subtest of the Inter-American Series Test of Reading, in Spanish, Level II and III for the Fourth Grade.

Test	Pre-test[1]			Post-test			Percentiles[2]	
	N	Mean	SD	N	Mean	SD	Pre-test	Post-test
Vocabulary	36	16.00	8.00	15	6.80	2.34	16	48
Speed of Comprehension	36	6.40	2.80	15	3.00	1.73	28	41
Level of Comprehension	36	13.40	6.40	15	6.26	3.26	16	38
Total Reading	36	35.80	13.90	15	16.06	5.58	15	40

[1] Test scores are raw scores.

[2] Percentiles are for fall semester and they belong to the Third grade from Puerto Rico in the pre-test, Level II and Fourth grade from Puerto Rico in the post-test, Level III.

The following information is drawn from this table:

- Children at the end of the third grade scored well below the norms in every subtest. Only between 16 and 28 percent of the third grade children in Puerto Rico scored lower than the mean score for

our sample. It must be added, again, that the normed sample was tested during the fall and children in school C.E.S. 235 during spring, thus making the differences more striking.

- There were differences between subtests on the size of the gap between percentile mean scores and the norms at the time of pre-testing. Children scored lower in Vocabulary and Level of Comprehension than in Speed of Comprehension; the difference was 12 percentile points.

- It must be said that the mean score of 6.40 in Speed of Comprehension is a little unreliable. The number of items in the subtest is 30 and the number of choices for each item is four; a score of seven can be obtained by guessing alone and, therefore, there is the probability that the mean score is the result of guessing.

- The sample finished third grade being a heterogeneous group according to their scores in Spanish Vocabulary and Level of Comprehension, as indicated by the relatively large size of the standard deviations; a wide variety of scores occurred in these subtests. The sample was a more homogeneous group according to the Speed of Comprehension scores; this indicates that most of the students scored below the score they could obtain by guessing alone.

- There was a positive improvement in every subtest during eleven months of school. Children scored higher at the end of the fourth grade than they scored the previous year.

- Students scored, however, below the norms for fourth grade children in Puerto Rico. None of the mean scores reached the 50th percentile of the norms.

- In order to understand the post-test scores we must keep in mind that all the post-test mean scores are well below the probable score that can be achieved by guessing alone. One finds himself a little uncomfortable in drawing conclusions from these scores.

f) Correlation Between the Subtests

Table 46 reports the coefficients of correlation between the three subtests for school C.E.S. 235 children and for the normed sample. From this table we have the following information.

Table 46: Product-Moment Correlation Coefficients Between Vocabulary, Speed and Level of Comprehension on the Inter-American Series Test of Reading in Spanish, Level III, for the Fourth Grade.

	C.E.S. 235	National Sample
Vocabulary vs. Speed of Comprehension	.30	.34
Vocabulary vs. Level of Comprehension	.03	.60
Speed of Comprehension vs. Level of Comprehension	.17	.62

- There was a low correlation between scores in Vocabulary and scores in Speed and Level of Comprehension. About nine percent of the scores in Speed of Comprehension could be explained by the scores in Vocabulary; the relationship between these two subtests had the highest coefficient. This means that less prediction can be obtained with the other coefficients.

- All the product-moment correlation coefficients in school C.E.S. 235 were lower than in the national sample. Less variance in a subtest could be predicted from scores in other subtests in our sample than in the national sample.

- The most important finding of this table is the fact that all the coefficients are low and, in one case, very close to zero. Each time we discussed the mean scores for each subtest we emphasized that the scores for the post-test were very low and that they could be the result of chance alone. In Table 46 we have information that supports this suspicion. If two scores are the product of chance alone they should not be correlated, and we can say with some probability of being right that the coefficients of correlation in school C.E.S. 235 were low because the scores were obtained mainly by guessing.

1) Test of Reading in English

a) Vocabulary

Table 47 contains the information about the mean and standard deviations for the pre-test, post-test and the corresponding percentile ranks for the English Vocabulary subtest.

Table 47: Vocabulary: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English, Level II and III for the Fourth Grade.

	Pre-test[1]	Post-test	Percentiles[2]		
			Pre-test	Post-test	Gain
N	24	39			
Mean	19.30	7.38	11	35	24
Standard Deviation	7.60	4.33			

[1]Test scores are raw scores.

[2]Percentiles are for:
Third grade in the pre-test, Level II
Fourth grade in the post-test, Level III.

From this table we can draw the following information:

- Children in school C.E.S. 235, as indicated by pre-test data, finished the third grade with a mean score in English Vocabulary that was well below the norms for native English speaking third grade children.

- About 89 percent of the third grade children from a national sample scored higher than the mean score for our sample. The result, however, was expected; cultural and language differences between Spanish speaking children and native English speaking children were expected to be reflected in vocabulary scores.

- There was a great variety of scores in English Vocabulary at the end of third grade as indicated by the size of the standard deviation. A range of more than 30 points is needed to include 96 percent of the sample scores. The possible range is 40. There were large differences among students in their knowledge of English words as indicated by the wide range of scores.

- Students in school C.E.S. 235 finished the fourth grade still below the norms as indicated by the post-test scores. About 65 percent of the fourth grade children from the norms scored as high or higher than the mean score for our sample.

- The sample became a more homogeneous group during the year as indicated by the relatively small standard deviation for the post-test. A range of about 16 points was needed to include 96 percent of the scores from a possible range of 45. This finding means that most of the students had a score below 16, which can be considered as a very low score.

- There was positive improvement during eleven months of school as indicated by a gain of 24 percentile points between pre and post-test.

- A note of caution must be added. A mean score of seven or even nine can be obtained simply by guessing. In a test of 45 items with five possible answers for each item, it is possible to get a score of nine without knowing a word of English and by guessing all the answers. If, in addition, a score of nine is equivalent to the 55th percentile ranks of the normed population, it means that the test is very difficult for the students, only 45 percent get a score as high or higher than the possible score they could obtain by guessing alone.

b) Speed of Comprehension

Table 48 contains the pertinent information about the Speed of Comprehension subtest. From this table we conclude:

Table 48: Speed of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English, Level II and III for the Fourth Grade.

	Pre-test[1]	Post-test	Percentiles[2]		
			Pre-test	Post-test	Gain
N	24	39			
Mean	8.20	4.87	15	34	19
Standard Deviation	4.70	3.02			

[1]Test scores are raw scores.

[2]Percentiles are for:
Third grade in the pre-test, Level II
Fourth grade in the post-test, Level III.

- Children in school C.E.S. 235, as indicated by the pre-test data, finished the third grade with a mean score in Speed of Comprehension of English that was well below the norms for native English speaking third grade children.

- About 85 percent of third grade children in the normed sample scored higher than the mean score for our sample. The result, again, was expected because of cultural and language differences between Spanish

speaking children and children from the national sample.

- There was a variety of scores in Speed of Comprehension of English at the end of the third grade as indicated by the size of the standard deviation. A range of about 19 points is needed to include 96 percent of the sample scores. The possible range is 40.

- Students in school C.E.S. 235 finished the fourth grade still below the norms as indicated by the post-test scores. About 66 percent of the fourth grade children from the normed sample scored as high or higher than the mean score for our sample.

- The sample became a more homogeneous group during the year as indicated by the relatively smaller standard deviation for the post-test. A range of about 12 points was needed to include 96 percent of the scores from a possible range of 45. This finding means that most of the students had a score below 11 which can be considered as a very low score.

- There was a positive improvement during eleven months of school as indicated by a gain of 19 percentile points between pre and post-test.

- Most of the children scored very low. The number of items in the test is 30, the choices for each item five; a score of six can be obtained by guessing all the answers. The mean score in the post-test of 4.87, therfore, can be the effect of a language ability or simply chance effect.

c) Level of Comprehension

From the information in Table 49 about the Level of Comprehension subtest we conclude that:

Table 49: Level of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English, Level II and III for the Fourth Grade.

	Pre-test[1]	Post-test	Percentiles[2]		
			Pre-test	Post-test	Gain
N	24	39			
Mean	15.90	8.15	3	26	23
Standard Deviation	7.40	5.35			

[1]Test scores are raw scores.

[2]Percentiles are for:
Third grade in the pre-test, Level II
Fourth grade in the post-test, Level III.

- Children in school C.E.S. 235, as indicated by the pre-test data, finished the third grade with a mean score in English Level of Comprehension that was well below the norms for the national sample.

- About 97 percent of the third grade children in the normed sample scored higher than the mean score for our sample. The result, again, was expected because of cultural and language differences between Spanish speaking children living in the United States and the national sample.

- There was a great variety of scores in Level of Comprehension of English at the end of third grade as indicated by the size of the standard deviation. A range of more than 29 points is needed to include 96 percent of the sample scores. The possible range is 40. There were large differences among students in their Level of Comprehension of English as indicated by the wide range of scores.

- Students in school C.E.S. 235 finished the fourth grade still below the norms as indicated by the post-test scores. About 74 percent of the fourth grade children in the normed sample scored as high or higher than the mean score for our sample.

- The sample became a more homogeneous group during the year as indicated by the relatively small standard deviation for the post-test. A range of about 21 points was needed to include 96 percent of the scores from a possible range of 50. This finding means that most of the students had a score below 18, which can be considered as very low.

- There was a positive improvement during eleven months of school as indicated by a gain of 23 percentile points between pre and post-test.

- A note of caution must be added once more. A mean score of eight or even ten can be obtained simply by guessing. In a test of 50 items, with five possible answers for each item, it is possible to get a score of ten without knowing a word of English and by guessing all the answers.

d) Total Reading

From Table 50 which contains the information about Total Reading data, we conclude that:

Table 50: Total Reading: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English, Level II and III for the Fourth Grade.

	Pre-test[1]	Post-test	Percentiles[2]		
			Pre-test	Post-test	Gain
N	24	39			
Mean	41.30	20.41	3	25	22
Standard Deviation	18.00	11.51			

[1]Test scores are raw scores.

[2]Percentiles are for:
Third grade in the pre-test, Level II
Fourth grade in the post-test, Level III.

- Children in school C.E.S. 235, as indicated by the pre-test data, finished the third grade with a mean score in Total Reading in English that was well below the norms for third grade children from the normed population.

- About 97 percent of third grade children in the national sample scored higher than the mean score for our sample. The result, again, was expected because of cultural and language differences between Spanish speaking children and native English speaking children.

- There was a great variety of scores at the end of the third grade as indicated by the size of the standard deviation. A range of more than 72 points is needed to include 96 percent of the sample scores.

The possible range is 125. There were large differences among students as indicated by the wide range of scores.

- Students in school C.E.S. 235 finished the fourth grade still below the norms as indicated by the post-test scores. About 75 percent of the fourth grade children in the normed sample scored as high or higher than the mean score for our sample.

- The sample became a more homogeneous group during the year as indicated by the relatively smaller standard deviation for the post-test. A range of about 46 points was needed to include 96 percent of the scores from a possible range of 125. This finding means that most of the students had a score below 42, which can be considered as a low score.

- There was a positive improvement during eleven months of school as indicated by a gain of 22 percentile points between pre and post-test.

- The mean score is below the possible score students can get by guessing alone which can be as high as 25.

e) Comparison of Subtests

Table 51 has the information for each subtest of the Test of Reading in English. The information is reported in a form which is similar to the form used in previous tables.

Table 51: Means and Standard Deviations for the Pre-test and Post-test, Percentile Ranks of each Subtest of the Inter-American Series Test of Reading, in English, Level II and III for the Fourth Grade.

Test	Pre-test[1]			Post-test			Percentiles[2]	
	N	Mean	SD	N	Mean	SD	Pre-test	Post-test
Vocabulary	24	19.30	7.60	39	7.38	4.33	11	35
Speed of Comprehension	24	8.20	4.70	39	4.87	3.02	15	34
Level of Comprehension	24	15.90	7.40	39	8.15	5.35	3	26
Total Reading	24	41.30	18.00	39	20.41	11.51	3	25

[1]Test scores are raw scores.

[2]Percentiles are for third grade in the pre-test, Level II, and fourth grade in the post-test, Level III.

The following information is drawn from this table:

- Children at the end of the third grade scored well below the norms in every subtest. Only between three and fifteen percent of third grade children from the national sample scored lower than the mean score for our sample.

- There were differences between subtests on the size of the gap between percentile mean scores and the norms at the time of pre-testing. Children scored lower in Vocabulary and Level of Comprehension than in Speed of Comprehension; the difference was between twelve and four percentile points.

- It must be said that the mean score of 4.87 in Speed of Comprehension is somewhat unreliable. The number of items is five; a score of six can be obtained by guessing alone. Therefore, there is the probability that the mean score is the result of guessing.

- The sample finished third grade being a heterogeneous group according to their scores in English Vocabulary and Level of Comprehension as indicated by the relatively large size of the standard deviations; a wide variety of scores occurred in these subtests. The sample was a more homogeneous group according to the Speed of Comprehension scores; this indicates that most of the students scored below the score they could obtain by guessing alone.

- There was a positive improvement in every subtest during eleven months of school. Children scored higher at the end of fourth grade than they scored the previous year.

- Students scored, however, below the norms for fourth grade children from a national sample. None of the mean scores reached the 50th percentile of the norms.

- In order to understand the post-test scores we must keep in mind that all the post-test mean scores are well below the probable score that can be achieved by guessing alone. One find himself a little uncomfortable in drawing conclusions from these scores.

f) Correlation Between Subtests

Table 52 reports the correlations between the subtests for school C.E.S. 235 and for the national sample.

Table 52: Product-Moment Correlation Coefficients Between Vocabulary and Speed and Level of Comprehension on the Inter-American Series Test of Reading in English, Level III for the Fourth Grade.

	C.E.S. 235	National Sample
Vocabulary vs. Speed of Comprehension	.69	.80
Vocabulary vs. Level of Comprehension	.68	.76
Speed of Comprehension vs. Level of Comprehension	.78	.80

This table indicates that:

- All the correlations between the subtests were substantial both in school C.E.S. 235 and in the national sample. This means that changes in one of the subtests are followed by similar changes in the other subtests. The amount of change in a subtest that can be explained by changes in other subtests is 46 percent between Vocabulary and Level of Comprehension, 48 percent between Vocabulary and Speed of Comprehension and 61 percent between Speed and Level of Comprehension.

- The coefficients are substantially lower in school C.E.S. 235 than in the national sample. The different reading skills are more interrelated in the national sample than in school C.E.S. 235 although the relationships are substantial in both samples.

c. Tests of Basic Language Competence

Tests of Language Competence in Spanish and English were given to all children who were judged by the teachers to be minimally competent in Spanish or English. Only subtests 1 and 2 were given. Different forms were used in Spanish and in English.

Since these tests have not been normed on a national sample, percentile ranks are not included here. The only possible comparison is to a perfect total score which is reported at the bottom of each table.

The test of Language Competence was designed to measure a child's ability to understand Spanish and English. Children who speak one or both languages as natives are expected to make near perfect scores on the subtests.

The subtest 1, Oral Vocabulary, presents words to the child from which the child must identify the corresponding picture. The subtest 2, Phonemic Discrimination, measures phonemic discrimination at word and phrase level. The test requires the child to discriminate whether two words sound the same or different.

1) Oral Vocabulary

Table 53 has the pertinent information about the Oral Vocabulary in Spanish and in English. From the data in this table we may conclude that:

Table 53: Oral Vocabulary: Means and Standard Deviations on Spanish and English Version of the Language Competence Test for the Fourth Grade.

Language	Pre-test[1]			Post-test			Mean Gain	t-test[2] Values
	N	Mean	SD	N	Mean	SD		
Spanish	40	16.00	8.70	29	20.17	4.69	4.17	2.34*
English	40	22.60	2.20	29	23.58	.86	.98	2.27*

[1]Test scores are raw scores.

[2]*Significant beyond the .05 level.

Note: Perfect score = 24.

- Children from school C.E.S. 235 scored very low in Spanish in the pre-test. Children finished the third grade with a knowledge of Spanish words that was less than perfect.

- Students formed heterogeneous groups according to their scores in Spanish in the pre-test. The highest possible range is 24 points but a range of more than 35 is needed in order to include only 96 percent of the scores in Spanish. This means that the sample was made up of two different groups, one with high scores and the other with very low scores, with few students scoring in-between.

- Children had practically a perfect mean score in English. They finished third grade with a perfect knowledge of simple English words.

- Students formed a relatively homogeneous group according to their scores at the end of the third grade. Very few students scored below perfect score.

- Students showed a significant improvement beyond the .05 level as indicated by the t-test values both in Spanish and in English during eleven months of school. The mean gain between pre and post-test was large enough to be beyond expectations of changes by chance alone.

- Children as a group scored an almost perfect score in English at the end of the fourth grade, and very few students scored below the perfect score as indicated by the small standard deviation. The mean score in Spanish was not near a perfect score; some students have problems with simple words in Spanish even at the end of the fourth grade.

- Children became a more homogeneous group both in Spanish and in English during the school year as indicated by the substantial decrease in the standard deviation.

2) Phonemic Discrimination

Table 54 reports the pertinent information about the subtest Phonemic Distrimination in Spanish and in English. From this table we have the following information:

Table 54: Phonemic Discrimination: Means and Standard Deviations in Spanish and English Version of the Language Competence Test for the Fourth Grade.

Language	Pre-test[1]			Post-test			Mean Gain	t-test[2] Values
	N	Mean	SD	N	Mean	SD		
Spanish	23	32.30	8.00	63	20.00	2.75	-12.30	10.62*
English	36	25.50	8.90	32	30.18	7.53	4.68	2.33*

[1]Test scores are raw scores.

[2]*Significant beyond the .05 level.

Note: Perfect score = 40.

- Children from school C.E.S. 235 scored low in both Spanish and English in the pre-test. Their mean score at the end of the third grade was less than perfect; their mastery of Spanish and English sounds was less than what is expected from children that speak either language as a native language. Students from our sample had a language disadvantage in relation to both Spanish and English speaking students.

- Children had a better discrimination of Spanish than English sounds as shown by the difference in mean score for the pre-test.

- Students were a heterogeneous group according to their scores in Spanish and in English as indicated by the size of the standard deviation for the pre-test. There was a wide range of scores in both languages.

- During eleven months of school, children decreased their mean score in Spanish, but they increased it in English as shown by the mean gains. Both changes were significant beyond the .05 level. The positive change they showed in English and the negative change in Spanish were changes indicating a deterioration of their ability to discriminate sounds in Spanish and an improvement in their ability to discriminate sounds in English.

- There was a substantial decrease in group heterogeneity in Spanish scores; students, as a group, sharply reduced the differences among themselves. In English there was a decrease in group heterogeneity but it was not as large as in Spanish. There was a wide range of English scores but the variety was less than the previous year.

7. Fifth Grade Achievement

Fifth grade children were given the Metropolitan Achievement Test, Reading, the Inter-American Series Test of Reading and the Basic Language Competence Test. Children were pre-tested at the end of fourth grade and post-tested at the end of the fifth grade using the same test level in both situations.

a. MAT-Reading

Children were tested in Word Knowledge and Reading subtests from the MAT, Elementary Level.

The 50 items of the Word Knowledge subtest measure extent of pupils' reading vocabulary. Items require pupils to identify synonyms, antonyms, or word classification. Items range from primary level to junior high level in difficulty.

The 45 items of the Reading subtest measure pupils' ability to comprehend written material. Pupils read a paragraph and then answer questions about it. Items cover comprehending literal meanings of passages, drawing inferences from the material, identifying main ideas, and determining word meanings from context.

1) Word Knowledge

The information related to the Word Knowledge subtest is reported in Table 55. This table contains information about the number of students and mean grade equivalent on the pre-test, post-test, gain between pre and post-test and difference between predicted and actual post-test.

Table 55: Word Knowledge: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Fifth Grade.

	Pre-test[1]	Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2]	
						Mean Gain	Difference
N	47	45					
Mean	2.90	3.94	1.04	3.45	.49	3.98*	1.71*
Standard Deviation	.96	1.50		1.24			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 4.80 at the time of the pre-test and 5.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored well below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, however, was expected given the special characteristics of school C.E.S. 235.

- The mean grade equivalent was 1 year, 9 months below the national level. This means that children in the program finished fourth grade with achievement scores at about the same level as the national norms for second grade.

- There were differences among the students in their Word Knowledge scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 1 year, 8 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

The column labeled "Mean Gain" merely indicates the difference between the pre-test and post-test mean scores. From this column we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 10 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 1 month during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school, the students were 1 month lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished the year of school being a heterogeneous group and they finished with a greater degree of heterogeneity than they had the previous year.

c) Program Impact

The column labeled "Predicted Post-test" indicates a statistical calculation based on pre-test scores. Each student's achievement thus far in school is divided by the number of months he has been in school (excepting kindergarten) and from this monthly mean gain the predicted post-test scores can be determined by simply multiplying the monthly mean gain by the number of months in the school between pre and post-testing and adding the result to the pre-test scores. The reason for this statistical manipulation is to determine whether the children gain more during the period in the program than they would have if they had continued making progress at the same rate they have thus far in their school career.

It should be noted that the predicted post-test scores are a direct reflection of the pre-test scores. If pre-test scores are low, a sample has gained very little up to that point and would be expected to gain little in the future without some special treatment. None of the school samples were expected to gain to their grade level if they were not at that level when they were pre-tested.

The column labeled "Difference" is the most important of the chart. It indicates the difference between what the children were expected to achieve and what they did achieve. From that column, the following conclusions can be drawn:

- The difference between actual and predicted post-test scores was positive. The children's mean score at the end of the year was higher than the mean expectations based on their scores at the end of the previous year. The school's impact was beneficial beyond expectations for the students as a group; during the year, students improved at a rate which was greater than the monthly rate of improvement up to that time .

- The mean difference between predicted and actual post-test was significant beyond the .05 level as indicated by the t-test value. The positive improvement of about 5 months beyond expectations that occurred in the students as a sample was large enough to be significant beyond changes expected by chance alone. We can say that the improvement was a real improvement in the ability measured by the test and it can not be explained by chance alone.

2) Reading

The information about the Reading subtest is reported in Table 56. The information is reported in a form similar to the form used in the previous table.

Table 56: Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Fifth Grade.

	Pre-test[1]	Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2] Mean Gain	Difference
N	47	45					
Mean	2.80	3.82	1.02	3.32	.50	3.80*	1.69*
Standard Deviation	1.04	1.50		1.34			

[1]The test scores are grade equivalents.

[2]* Significant beyond the .05 level.

NOTE: The National Grade Level was 4.80 at the time of the pre-test and 5.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, again, was expected given the special characteristics of school C.E.S. 235.

- The mean grade equivalent was 2 years below the national level. This means that children in the program finished fourth grade with achievement scores at the level of the national norms for second grade.

- There were differences among the students in their Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 2 years, 1 month above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "Mean Gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 10 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 1 month during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school, the students were 1 month lower in relation to the national sample than they were a year earlier. Students improved during the year but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished the year of school being a heterogeneous group and they finished with a greater degree of heterogeneity than they had the previous year.

c) Program Impact

From the column labeled "Difference" we draw the following conclusions:

- The difference between actual and predicted post-test scores was positive. The children's mean score at the end of the year was higher than the mean expectation based on their scores at the end of the previous year. The school's impact was beneficial for the students as a group; during the year, students improved at a rate which was greater than the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test was significant beyond the .05 level as indicated by the t-test value. The positive improvement of about 5 months beyond expectations that occurred in the students as a sample was large enough to be significant beyond changes expected by chance alone. We can say that the improvement was a real improvement in the ability measured by the test and it can not be explained by chance alone.

3) Total Reading

Table 57 has the information about Total Reading. The information is reported in a form similar to the form used in previous tables.

Table 57: Total Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Fifth Grade.

	Pre-test[1]	Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2]	
						Mean Gain	Difference
N	47	45					
Mean	2.70	3.81	1.11	3.19	.62	4.26*	2.16*
Standard Deviation	.99	1.47		1.28			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 4.80 at the time of the pre-test and 5.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result,

again, was expected given the special characteristics of school C.E.S. 235.

- The mean grade equivalent was 2 years, 1 month below the national level. This means that children in the program finished third grade with achievement scores lower than the national norms for second grade.

- There were differences among the students in their Total Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample, a range of 2 years above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "Mean Gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 11 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level remained the same during the school period. Children, after a year in the program, were still behind the national norms; students improved during the year but not more than the national sample and, therefore, the gap between mean grade equivalent and national level remained the same at the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished the year of school being a heterogeneous group and they finished with a greater degree of heterogeneity than they had a year before. Children's improvement was not similar throughout the range of their scores at the end of the previous year as shown by the increased size of the standard deviation at the end of the year.

c) Program Impact

From the column labeled "Difference" we conclude that:

- The difference between actual and predicted post-test scores was positive. The children's mean score at the end of the year was higher than the mean expectation based on their scores at the end of the previous year. The school's impact was beneficial beyond expectations for the students as a group; during the year students improved at a rate which was greater than the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test was significant beyond the .05 level as indicated by the t-test value. The positive improvement of about 6 months beyond expectations that occurred in the students as a sample was large enough to be significant beyond changes expected by chance alone. We can say that the improvement was a real improvement in the ability measured by the test.

5) Comparison of Subtests

Table 58 includes information about each subtest and the total test. Information is arranged in a form similar to the form used in the previous tables. From this table we have the following additional information:

Table 58: Means and Standard Deviation for the Pre-test, Actual Post-test, Predicted Post-test and t-test Values for the Difference Between Pre-test, Actual Post-test and Predicted Post-test on Each Subtest of the MAT-Reading for the Fifth Grade.

Test[4]	Pre-test[1]			Actual Post-test			Mean Gain	Predicted Post-test		Difference	t-test Values	
	N	Mean	SD	N	Mean	SD		Mean	SD		2	3
Word Knowledge	47	2.90	.96	45	3.94	1.50	1.04	3.45	1.24	.49	3.98*	1.71*
Reading	47	2.80	1.04	45	3.82	1.50	1.02	3.32	1.34	.50	3.80*	1.69*
Total Reading	47	2.70	.99	45	3.81	1.47	1.11	3.19	1.28	.62	4.26*	2.16*

[1]The test scores are grade equivalents.

[2]t-test values for the difference between pre and actual post-test mean-scores.

[3]t-test values for the difference between actual and predicted post-test mean scores.

*t-test value significant beyond the .05 level.

[4]Correlation between subtests: r= .88.

NOTE: The National Grade Level was 4.80 at the time of pre-test and 5.90 at the time of the post-test.

a) Pre-program Level

From pre-test scores it can be said that:

- Children scored below national norms in both subtests, Word Knowledge, and Reading.
- Children finished fourth grade about 2 years below national norms in Word Knowledge and in Reading.
- The gap between grade equivalent and national norms was very similar in both subtests; students scored 1 month lower in Reading than in Word Knowledge.
- Students were heterogeneous groups in every subtest according to their

scores. The degree of heterogeneity was similar in both subtests.

b) Pre-post Test Change

From the column labeled "Mean Gain" the following conclusions can be drawn:

- There was a positive gain in grade equivalent between pre and post-test. Children, after eleven months in school, scored higher than when they finished the year before. They improved in both subtests.

- Children improved the same in Word Knowledge and in Reading; the difference between mean gains was negligible.

- The gap between mean grade equivalent and national norms increased in both Word Knowledge and Reading. During eleven months of school, children in the sample improved only 10 months in Word Knowledge and in Reading, thus increasing the gap by 1 month. The gap, however, remained the same in Total Reading.

- There was a substantial increase in group heterogeneity in both subtests as indicated by the size of the standard deviations for the pre and post-test.

c) Program Impact

From column labeled "Difference" it can be concluded that:

- There was a positive difference in Word Knowledge and Reading between predicted and actual post-test.

- The program's impact was beneficial in Word Knowledge, Reading and Total Reading. Students scored higher than was expected of them with expectations based on their past improvement.

- The changes beyond expectations were significant at the .05 level. The changes in Word Knowledge, Reading and Total Reading were real changes in the abilities measured by the test, and they can not be a result of chance alone.

- The correlation between subtests was high. With a coefficient of .88, about 77 percent of the raw scores in one of the subtests can be predicted from the scores on the other subtest. Changes in one subtest are followed by substantial and similar changes in the other subtest.

b. Inter-American Series Test of Reading in English

The Test of Reading in English was given to all children, either Spanish or English dominant. Level II was used with fifth grade children. Form CE was used in the pre-test and in the post-test. The children were not tested in Spanish.

Caution should be used against attributing too much significance to percentile ranks as this test was normed on small, specific groups of children. In Level III, in the Vocabulary test the stem is a definition, a descriptive phrase, or a word suggesting the response; the task for the children is to find the word among different choices. In Speed of Comprehension the task is to find a word omitted from the sentence in the stem; the items are arranged in cycles of similar difficulty, and the test is highly speeded. The Level of Comprehension is tested by questions based on paragraphs of generally increasing difficulty.

Careful note must be made that the post-test percentiles discussed in the following tables are both related to the mean raw scores for the post-test. The post-test, as stated earlier, was given to the children during the opening semester of their fifth grade. With children scoring well below the national norms, we confront two problems: 1) did they improve at all during the program year, and 2) if they made progress, did they improve as much as the national sample. In order to answer the first question, the mean raw score of the post-test was converted into percentile ranking at the same level as the pre-test (i.e., fourth grade). To answer the second question, the mean raw score must be compared with that of the fifth grade national sample.

1) Vocabulary

Table 60 contains the information about the means and standard deviations for the pre and post-test, the correspondent percentile ranks for the means and the t-test value for the mean gain and the Vocabulary subtest of the Reading test in English.

Table 60: Vocabulary: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level III for the Fifth Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles			t-test Values
				Pre-test 4th Grade	Post-test 4th Grade	Post-test 5th Grade	
N	26	42					
Mean	12.40	12.88	.48	62	66	43	.27[2]
Standard Deviation	6.20	7.41					

[1]Test scores are raw scores.

[2]t-test value = .27 which was not significant at the .05 level.

From this table we have the following information:

- Children in school C.E.S. 235 finished fourth grade with a mean score in English Vocabulary **that was above the 50th percentile for native English speaking** fourth grade children, as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of 62 percent of second grade students from a national **sample.** This result was not expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the English Vocabulary score.

- There were a variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 25 points would be needed in order to include 96 percent of the scores in the sample; 45 is the possible range of scores.

- During eleven months of school, children improved in their English Vocabulary scores. The mean score increased by 4 percentile points.

- Children in C.E.S. 235 did not improve as much as the national sample. Children from our sample finished fourth grade above the level of a national fourth grade sample; but at the end of the fifth grade they were 7 percentile points below the national fifth grade level thus showing an improvement substantially smaller than what is expected from a national sample.

- The mean gain of 4 percentile points was not significant beyond the .05 level as indicated by the t-test value. The improvement in English Vocabulary was smaller than what could be expected by chance alone; it was not a real improvement in the trait measured by the test.

- There was a small increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores in English Vocabulary at the end of fourth grade and they had a greater degree of heterogeneity at the end of fifth grade.

2) Speed of Comprehension

Table 61 contains the pertinent information related to the Speed of Comprehension subtest. From this table we conclude that:

Table 61: Speed of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level III for the Fifth Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles			t-test[2] Values
				Pre-test 4th Grade	Post-test 4th Grade	Post-test 5th Grade	
N	27	42					
Mean	6.50	8.16	1.66	56	67	43	1.26
Standard Deviation	4.00	6.06					

[1]Test scores are raw scores.

[2]t-test value = 1.26 which was not significant at the .05 level.

- Children in school C.E.S. 235 finished fourth grade with a mean score in Speed of Comprehension that was above the 50th percentile for native English speaking fourth grade children, as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of 56 percent of fourth grade students from at national sample. This result was not expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the Speed of Comprehension score.

- There were a variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 16 points would be needed in order to include 96 percent of the scores in the sample; 30 is the possible range of scores.

- During eleven months of school, children improved in their Speed of Comprehension scores. The mean score increased by 11 percentile points.

- Children in C.E.S. 235 did not improve as much as the national sample. Children from our sample finished fourth grade above the level of a national fourth grade sample; but at the end of the fifth grade they were 7 percentile points below the national fifth grade level thus showing an improvement substantially smaller than what is expected from a national sample.

- The mean gain of 11 percentile points was not significant at the .05 level as indicated by the t-test value. The improvement in Speed of Comprehension was smaller than what could be expected by chance alone; it was not a real improvement in the trait measured by the test.

- There was an increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores at the end of fourth grade and they had a greater degree of heterogeneity at the end of fifth grade.

3) Level of Comprehension

Table 62 has the pertinent information about the Level of Comprehension subtest. The following conclusions are drawn from this table:

Table 62: Level of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, In English Level III for the Fifth Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles			t-test[2] Values
				Pre-test 4th Grade	Post-test 4th Grade	Post-test 5th Grade	
N	28	42					
Mean	8.90	14.76	5.86	30	59	36	3.82
Standard Deviation	4.20	7.34					

[1] Test scores are raw scores.

[2] t-test value = 3.82 which was significant beyond the .05 level.

- The pre-test mean score should be interpreted with great care, because it is very unreliable. In a test of 50 items with 5 choices for each item, as it is the case in Level of Comprehension subtest, it is quite possible to get a score of 9 simply by quessing. The mean score is below this guessing score.

- Children in school C.E.S. 235 finished fourth grade with a mean score in Level of Comprehension that was below the 50th percentile for native English speaking fourth grade children, as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of only 30 percent of fourth grade students from a national sample. This result was expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the Level of Comprehension score.

- There were a variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 17 points would be needed in order to include 96 percent of the scores in the sample; 50 is the possible range of scores.

- During eleven months of school, children improved in their Level of Comprehension scores. The mean score increased by 29 percentile points.

- Children improved more than the national sample. At the end of fourth grade our sample children had a mean score that was at the 30th percentile of a fourth grade national sample. At the end of the fifth grade their mean score was at the 36th percentile of the fifth grade national sample. They reduced the gap with a national sample by 6 percentile points.

- The mean gain of 29 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in Level of Comprehension was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a substantial increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores at the end of fourth grade and they had a greater degree of heterogeneity at the end of fifth grade.

4) Total Reading

Table 63 has the information related to the Total Reading test. The following conclusions are drawn from this table:

Table 63: Total Reading: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level III for the Fifth Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles			t-test[2] Values
				Pre-test 4th Grade	Post-test 4th Grade	Post-test 5th Grade	
N	26	42					
Mean	27.00	35.57	8.57	41	62	38	2.01
Standard Deviation	12.60	19.26					

[1] Test scores are raw scores.

[2] t-test value = 2.01 which was significant beyond the .05 level.

- Children in school C.E.S. 235 finished fourth grade with a mean score in Total Reading that was below the 50th percentile for native English speaking fourth grade children, as reported by the pre-test scores. The mean score for our sample is as high or higher than only 41 percent of fourth grade students from a national sample. This result was expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the Total Reading scores.

- There were a great variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 50 points would be needed in order to include 96 percent of the scores in the sample; 45 is the possible range of scores.

- During eleven months of school, children improved in their Total Reading scores. The mean score increased by 21 percentile points.

- Children in C.E.S. 235 did not improve as much as the national sample. Children from our sample finished fourth grade at the 41st percentile level of a national fourth grade smaple; but at the end of the fifth grade they scored at the 38th percentile of the national fifth grade level thus showing an improvement smaller than what is expected from a national sample.

- The mean gain of 21 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in Total Reading was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a substantial increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores at the end of fourth grade and they had a greater degree of heterogeneity at the end of the fifth grade.

5) Comparison of Subtests

Table 64 includes the information related to each subtest and the total test. From this table we draw the following conclusions:

Table 64: Means and Standard Deviations for the Pre-test, Post-test, Percentile Ranks of Each Subtest of the Inter-American Series Test of Reading, in English Level III, for the Fifth Grade.

Test	Pre-test[1]			Post-test			Mean Gain	Percentiles			t-test[2] Values
	N	Mean	SD	N	Mean	SD		Pre-test 4th	Post-test 4th	Post-test 5th	
Vocabulary	26	12.40	6.20	42	12.88	7.41	.48	62	66	43	.27
Speed of Comprehension	27	6.50	4.00	42	8.16	6.06	1.66	56	67	43	1.26
Level of Comprehension	28	8.90	4.20	42	14.76	7.34	5.86	30	59	36	3.82*
Total Reading	26	27.00	12.60	42	35.57	19.26	8.57	41	62	38	2.01*

[1] Test scores are raw scores.

[2] *Significant beyond the .05 level.

- Children in school C.E.S. 235 finished fourth grade, as reported by the pre-test data, with a mean score in English Vocabulary and Speed of Comprehension, higher than the 50th percentile of the native English speaking fourth grade children. They scored lower than the national sample in Level of Comprehension and Total Reading.

- There was a variety of scores in every subtest as indicated by the size of the standard deviations. The group was a heterogeneous group and it had different degrees of heterogeneity on each subtest. The greatest range of scores occurred in Vocabulary.

- During eleven months of school, children improved in every subtest. The greatest improvement was in Level of Comprehension where the difference between pre and post-test mean scores was 29 percentile points.

- The mean gains were significant beyond the .05 level only in Level of Comprehension and Total Reading as indicated by the t-test values. The mean gains in Vocabulary and Speed of Comprehension were not beyond what could be expected by chance alone and, therefore, we can not say that the gains were real gains in the abilities measured by the subtests.

- Children improved less than a national sample in every subtest, excepting the Level of Comprehension subtest. At the end of fourth grade, school C.E.S. 235 children scored above the 50th percentile in Vocabulary and Speed of Comprehension, but at the end of the fifth grade they scored below the 50th percentile in both subtests; in Total Reading they were 9 percentile points behind the norms at the end of the fourth grade and 12 percentile points at the end of the fifth grade, thus increasing the gap by 3 percentile points. There was a small reduction of the gap, though, in Level of Comprehension; children's mean score reached only the 30th percentile in the fourth grade but it was at the 36th percentile in the fifth grade, thus reducing the gap by 6 percentile points.

6) Correlation Between Subtests

Table 65 reports the correlations between the three subtests for school CES 235 and for the national sample.

Table 65: Product-Moment Correlation Coefficients Between Vocabulary, Speed and Level of Comprehension on the Inter-American Series Test of Reading in English, Level III, for the Fifth Grade.

	C.E.S.235	National Sample
Vocabulary Vs. Speed of Comprehension	.81	.76
Vocabulary Vs. Level of Comprehension	.75	.76
Speed of Comprehension Vs. Level of Comprehension	.72	.79

From this table we have the following conclusions:

- The correlations between subtests were high in school C.E.S. 235. This means that changes in scores on one of the subtests were followed by substantial changes and in the same way on the other subtests. The amount of change in a subtest that can be explained by changes in other subtest are as follows: about 66 percent between Vocabulary and Speed of Comprehension, 56 percent between Vocabulary and Level of Comprehension, and 52 percent between Speed of Comprehension and Level of Comprehension.

- The correlation coefficients between Vocabulary and Speed of Comprehension were a little higher in school C.E.S. 235 than in the national sample. This means that improvement in English Vocabulary was more related in our sample than in a national sample to improvements in Speed of Comprehension. The opposite was also true; low scores in Speed of Comprehension were the result of low scores in Vocabulary. This connection was stronger in our sample than in a national sample.

- The correlation between Speed and Level of Comprehension and between Vocabulary and Level of Comprehension were a little lower in school C.E.S. 235 than in a national sample.

c. Tests of Basic Language Competence

Tests of Language Competence in Spanish and English were given to all children who were judged by the teachers to be minimally competent in Spanish or English. Only subtests 1 and 2 were given. Different forms were used in Spanish and in English.

Since these tests have not been normed on a national sample, percentile ranks are not included here. The only possible comparison is to a perfect total score, which are reported at the bottom of each table.

The test of Language Competence was designed to measure a child's ability to understand Spanish and English. Children who speak one or both languages as a native are expected to make near perfect scores on the subtests.

The subtest 1, Oral Vocabulary, presents words to the child from which the child must identify the corresponding picture. The subtest 2 measures phonemic

discrimination at word and phrase level; the test requires the child to discriminate whether two words sound the same or different.

1) Oral Vocabulary

Table 66 has the pertinent information about the Oral Vocabulary in Spanish and in English. From the data in this table we may conclude that:

Table 66: Oral Vocabulary: Means and Standard Deviations on Spanish and English Version of the Language Competence Test for the Fifth Grade.

Language	Pre-test[1]			Post-test			Mean Gain	t-test[2] Values
	N	Mean	SD	N	Mean	SD		
Spanish	21	23.60	.90	54	19.70	6.88	-3.90	2.58*
English	21	23.20	1.40	52	23.40	2.10	.20	.40

[1]Test scores are raw scores.

[2]*Significant beyond the .05 level.

NOTE: Perfect score = 24.

- Children from school C.E.S. 235 had practically a perfect mean score in Spanish and in English in the pre-test. Children finished fourth grade with a perfect knowledge of simple Spanish and English words.

- Students formed an homogeneous group according to their scores in Spanish and in English in the pre-test. The highest possible range is 24 and only a range of 5.6 is needed in order to include 96 percent of the scores in English where they had the greatest variability.

- Children did not change significantly in English Oral Vocabulary during eleven months of school. There was not too much room for change since most of them finished second grade with a perfect score. The positive mean change of .20 between pre and post-test was not significant and it is the result of chance alone.

- Students showed a loss in Spanish during the eleven months of school. The negative mean gain of 2.90 was large enough to be beyond changes expected by chance alone. The loss was in the ability measured by the test.

- Children became a more heterogeneous group in their Spanish scores during the school year as indicated by a substantial increase in the size of the standard deviation between the pre and post-testing. They became a little more heterogeneous as a group in English also.

2) Phonemic Discrimination

Table 36 reports the pertinent information about the subtest Phonemic Discrimination in Spanish and in English. From this table we have the following information:

Table 67: Phonemic Discrimination: Means and Standard Deviations on Spanish and English Version of the Language Competence Test for the Fifth Grade.

Language	Pre-test[1]			Post-test			Mean Gain	t-test[2] Values
	N	Mean	SD	N	Mean	SD		
Spanish	20	32.70	9.50	54	20.72	2.81	-11.98	8.41*
English	21	28.70	11.50	29	33.72	7.33	5.02	1.88*

[1]Test scores are raw scores.

[2]*Significant beyond the .05 level.

NOTE: Perfect score = 40.

- Children from school C.E.S. 235 scored low in Spanish and in English in the pre-test. Their mean score at the end of fourth grade was less than perfect; their mastery of Spanish and English sounds was less than expected from children that speak either language as a native language. Students from our sample had a language disadvantage in relation to both Spanish and English speaking students.

- Children had a better discrimination of Spanish than of English sounds as indicated by the difference in mean scores for the pre-test. The difference, though, was small.

- Students were a heterogeneous group according to their scores in Spanish and in English as indicated by the size of the standard deviation for the pre-test. There was a wide range of scores in both languages.

- During eleven months of school, children decreased their mean score in Spanish as indicated by the negative mean gain. The negative change was significant beyond the .05 level and it represented a loss in the ability measured by the test. The difference between pre and post-test was large enough to be beyond changes expected by chance alone.

- The mean gain in English was positive and significant beyond the .05 level. Children showed a real improvement in the ability of discriminating English phonemics.

- There was a substantial reduction in group heterogeneity in Spanish and English scores; students, as a group, reduced the differences among them.

8. Sixth Grade Achievement

Sixth grade students were given the Metropolitan Achievement Test Reading, the Inter-American Series Test of Reading and the Basic Language Competence Test. Children were pre-tested at the end of the fifth grade and post-tested at the end of the sixth grade using the same test level in both situations.

a. MAT-Reading

Children were tested in Word Knowledge and Reading subtests from the MAT, Intermediate Level.

The 50 items of the Word Knowledge subtest measure the extent of pupils' reading vocabulary. Items ask for synonyms, antonyms, or classification of words. Content covers words used for general information, social studies, humanities, science, and mathematics, without calling for knowledge of technical terms.

The 45 items of the Reading subtest measure pupils' ability to comprehend written material. Pupils read a paragraph, then answer questions about it. Items call for comprehending literally stated material, drawing inferences from a passage, identifying main ideas, and determining word meanings from context.

1) Word Knowledge

The information related to the Word Knowledge subtest is reported in Table 68. This table contains information about the number of students and mean grade equivalent on the pre-test, post-test, gain between pre and post-test and difference between predicted and actual post-test.

Table 68: Word Knowledge: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Sixth Grade.

	Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2]	
						Mean Gain	Difference
N	25	22					
Mean	3.60	4.93	1.33	4.20	.73	3.48*	1.72*
Standard Deviation	1.21	1.41		1.49			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 5.80 at the time of the pre-test and 6.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following.

- Children as a group scored well below the national noms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, though, was expected given the special characteristics of school C.E.S. 235.

- The mean grade equivalent was 2 years and 1 month below the national level. This means that children in the program finished fifth grade with achievement scores lower than the national norms for third grade.

- There were differences among the students in their Word Knowledge scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 2 years, 4 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

The column labeled "Mean Gain" merely indicates the difference between the pre-test and post-test mean scores. From this column we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 1 year and 3 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level decreased by 2 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 2 months higher in relation to the

national sample than they were a year earlier. Students improved during the year more than the national sample did and, therefore, the gap between mean grade equivalent and national level decreased by the end of the year.

- There was an increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished the school year being a heterogeneous group, and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

The column labeled "Predicted Post-test" indicates a statistical calculation based on pre-test scores. Each student's achievement thus far in school is divided by the number of months he has been in school (excepting kindergarten) and from this monthly mean gain the predicted post-test scores can be determined by simply multiplying the monthly mean gain by the number of months in the school between pre and post-testing and adding the result to the pre-test scores. The reason for this statistical manipulation is to determine whether the children gain more during the period in the program than they would have if they had continued making progress at the same rate they have thus far in their school career.

It should be noted that the predicted post-test scores are a direct reflection of the pre-test scores. If pre-test scores are low, a sample has gained very little up to that point and would be expected to gain little in the future without some special treatment. None of the school sample were expected to gain to their grade level if they were not at that level when they were pre-tested.

The column labeled "Difference" is the most important of the chart. It indicates the difference between what the children were expected to achieve and what they did achieve. From that column the following conclusions can be drawn:

- The difference between actual and predicted post-test scores was positive. The children's mean score at the end of the year was higher than the mean expectation based on their scores at the end of the previous year. The school's impact was beneficial beyond expectations for the students as a group; during the year students improved at a rate which was greater than the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test was significant beyond the .05 level as indicated by the t-test value. The positive improvement of about 7 months beyond expectations that occurred in the students as a sample was large enough to be significant beyond changes expected by chance alone. We can say that the improvement was a real improvement in the ability measured by the test and it can not be explained by chance alone.

2) Reading

The information related to the Reading subtest is reported in Table 69. The information is reported in a form similar to the form used in the previous table.

Table 69: Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Sixth Grade.

	Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2] Mean Gain	t-test Values[2] Difference
N	25	22					
Mean	3.80	4.77	.97	4.44	.33	2.28*	.70
Standard Deviation	1.29	1.62		1.59			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 5.80 at the time of the pre-test and 6.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored well below the national norms when they were pre-tested. They scored lower than expected of them if the conformed to the national sample. The result, though, was expected given the special characteristics of school C.E.S. 235.

- The mean grade equivalent was 2 years, 1 month below the national level. This means that children in the program finished fifth grade with achievement scores lower than the national norms for third grade.

- There were differences among the students in their Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 2 years and 6 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "Mean Gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored about 10 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level increased by 1 month during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 1 month lower in relation to the national sample than they were a year

earlier. Students improved during the year, but not as much as the national sample did and, therefore, the gap between mean grade equivalent and national level increased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviations for the pre and post-test. Students, as a group, finished the school year being a heterogeneous group, and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

From the column labeled "Difference" the following conclusions are drawn:

- The difference between actual and predicted post-test scores was positive. Children's mean score at the end of the year was higher than the mean expectation based on their scores at the end of the previous year. The school's impact was beneficial beyond expectations for the students as a group; during the year, students improved at a rate which was greater than the monthly rate of improvement up to that time.

- The mean difference between predicted and actual post-test, though, was not significant at the .05 level as indicated by the t-test value. The positive improvement of about 3 months beyond expectations that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the improvement was a real improvement in the ability measured by the test and it can be explained by chance alone.

3) Total Reading

The information related to the Total Reading subtest is reported in Table 70. The information is reported in a form similar to the form used in previous tables.

Table 70: Total Reading: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test, Mean Gain Between Pre and Post-test, Difference Between Predicted and Actual Post-test and t-test Values on the MAT for the Sixth Grade.

	Pre-test[1]	Actual Post-test	Mean Gain	Predicted Post-test	Difference	t-test Values[2]	
						Mean Gain	Difference
N	25	22					
Mean	3.60	4.92	1.32	4.20	.72	3.23*	1.60
Standard Deviation	1.24	1.56		1.52			

[1]The test scores are grade equivalents.

[2]*Significant beyond the .05 level.

NOTE: The National Grade Level was 5.80 at the time of the pre-test and 6.90 at the time of the post-test.

a) Pre-program Level

The data for the pre-test scores as indicated in the table show the following:

- Children as a group scored well below the national norms when they were pre-tested. They scored lower than expected of them if they conformed to the national sample. The result, though, was expected given the special characteristics of school C.E.S. 235.
- The mean grade equivalent was 1 year, 4 months below the national level. This means that children in the program finished fifth grade with achievement scores lower than the national norms for third grade.
- There were differences among the students in their Total Reading scores as indicated by the size of the standard deviation. In order to include 98 percent of the sample a range of 2 years, 5 months above and below the mean would be necessary. The sample was a heterogeneous group with a wide range of scores above and below the group mean.

b) Pre-post Test Change

From the column labeled "Mean Gain" we can see that:

- There was a positive gain in grade equivalent between pre and post-testing. Children, after eleven months in the school, scored 1 year and 3 months higher than their mean score a year earlier. The gain was significant at the .05 level as indicated by the t-test value and it was a real gain in the ability measured by the test.

- The gap between mean grade equivalent and national level decreased by 2 months during the school period. This means two things: 1) children, after a year in the program, were still behind the national norms; and 2) after a year in the school the students were 2 months higher in relation to the national sample than they were a year earlier. Students improved during the year more than the national sample did and, therefore, the gap between mean grade equivalent and national level decreased by the end of the year.

- There was a substantial increase in group heterogeneity as indicated by the size of the standard deviation for the pre and post-test. Students, as a group, finished the school year being a heterogeneous group, and they finished with a greater degree of heterogeneity than they had a year before.

c) Program Impact

From the column labeled "Difference" the following conclusions are drawn:

- The difference between actual and predicted post-test scores was positive. Children's mean score at the end of the year was higher than the mean expectation based on their scores at the end of the previous year. The school's impact was beneficial beyond expectations for the students as a group; during the year, students improved at a rate which was

greater than the montly rate of improvement up to that time.

- The mean difference between predicted and actual post-test though was not significant at the .05 level as indicated by the t-test value. The positive improvement of about 7 months beyond expectation that occurred in the students as a sample was not large enough to be significant beyond changes expected by chance alone. We can not say that the improvement was a real improvement in the ability measured by the test and it can be explained by chance alone.

4) Comparison of Subtests

Table 71 includes information about each subtest and the total test. Information is arranged in a form similar to the form used in the previous tables. From this table we have the following additional information:

Table 71: Means and Standard Deviations for the Pre-test, Actual Post-test, Predicted Post-test and t-test Values for the Difference Between Pre-test, Actual Post-test and Predicted Post-test on Each Subtest of the MAT-Reading for the Sixth Grade.

Test[4]	Pre-test[1]			Actual Post-test			Mean Gain	Predicted Post-test		Difference	t-test Value	
	N	Mean	SD	N	Mean	SD		Mean	SD		2	3
Word Knowledge	25	3.60	1.21	22	4.93	1.41	1.33	4.20	1.49	.73	3.48*	1.72*
Reading	25	3.80	1.29	22	4.77	1.62	.97	4.44	1.59	.33	2.28*	.70
Total Reading	25	3.60	1.24	22	4.92	1.56	1.32	4.20	1.52	.72	3.23*	1.60

[1]The test scores are grade equivalents.

[2]t-test values for the difference betwee pre and actual post-test mean scores.

[3]t-test values for the difference between actual and predicted post-test mean scores.

*t-test value significant beyond the .05 level.

[4]Correlation between subtests: r = .79.

NOTE: The National Grade Level was 5.80 at the time of the pre-test and 6.90 at the time of the post-test.

a) Pre-program Level

From pre-test scores it can be said that:

- Children scored below national norms in the two subtests, Word Knowledge and Reading.
- Children finished fifth grade 2 years and 2 months below national norms in Word Knowledge and 2 years and 1 month below national norms in Reading.
- The gap between grade equivalent and national norms was very similar in every subtest; students scored 2 months higher in Reading than in Word Knowledge.
- Students were a heterogeneous group in both subtest according to their scores. The degree of heterogeneity was similar in both subtests.

b) Pre-post Test Change

From the column labeled "Mean Gain:, the following conclusions can be drawn:

- There was a positive gain in grade equivalent between pre and post-test. Children, after eleven months in school, scored higher than when they finished the year before. They improved in the two subtests.
- Children improved less in Reading than they did in Word Knowledge, the difference between mean gains being about 3 months.
- The gap between mean grade equivalent and national norms increased in Reading but it decreased in Word Knowledge. During eleven months of school, children in the sample improved 10 months in Reading and they improved 13 months in Word Knowledge, thus increasing the gap by 1 month in Reading and reducing it by 2 months in Word Knowledge.
- There were substantial increases in group heterogeneity in both subtests as indicated by the size of the standard deviations for the pre and post-test.

c) Program Impact

From column labeled "Difference", it can be concluded that:

- There was a positive difference in Word Knowledge and Reading between predicted and actual post-test.
- The program's impact was beneficial both in Word Knowledge and Reading and on the Total Reading. Students scored higher than was expected of them, with expectations based on their past improvement.
- Changes beyond expectations were significant only in Word Knowledge but not in Reading or Total Reading. Changes beyond expectations in Reading and Total Reading do not represent real changes in the abilities measured by the test and they can be the result of chance alone. On the other hand, chance alone can not explain improvements beyond expectations in Word Knowledge and the changes represent real improvement in the knowledge of English words.
- The correlation between subtests was high. About 62 percent of the scores on one of the subtest can be predicted from the scores on the other. Improvements on one subtest are followed by improvements in the other subtest that are both substantial and corresponding.

d) A Further Analysis of Program's Impact

In Table 71 there is information in the column labeled "Difference" that deserves a little more explanation. A Difference of .73 for the Word Knowledge is barely significant, but the difference of .72, almost identical, for Total Reading is not significant. It is difficult to understand this result because both changes are substantial, i.e., larger than what is usually found in most of the programs. To improve 7 months beyond expectation during the school year is a great deal of improvement and everybody expects it to be not only significant but also highly significant. The resultant t-test values are, therefore, a little shocking.

The t-test is only a statistical manipulation which tries to measure changes that can be expected by chance alone. The amount of change expected by chance alone depends on two factors: 1) number of subjects in the sample and 2) variability among the sample.

The smaller the number of subjects in a sample, the greater the possibility that a change in the group's mean score is the result of chance alone or the effect of some factors unrelated to the program. With a small number of subjects it is highly improbable that the whole range of effect of every other factor is included; with a large sample, on the other hand, it is probable that positive and negative effects of other factors are included, thus making the mean effect for the other factors zero or near zero.

The greater the variability of a sample the higher is the probability that a change in the group's mean score is the result of chance alone. If there is a great variety of scores, few subjects have the same score and, therefore, there is a high probability of not including the whole range of effects of every other factor and there is also a high probability that the change is not due to the program.

Both probelms occurred in our sample of sixth graders to make it more difficult to find a significant difference. The sample was small and the variability among the subjects very high. This is the reason why changes as substantial as 7 months were not significant. The same changes with bigger samples or with more homogeneous samples would be highly significant.

Therefore, although the changes were remarkable, we could not prove that they were beyond expectations by chance alone, but it does not necessarily mean they were the effect of chance.

b. Inter-American Series Test of Reading

The Test of Reading in English was given to all children, either Spanish or English dominant. The Test of Reading in Spanish was not given to sixth grade students. Level III was used with this grade. Form CE was used in the pre and in the post-test.

Caution should be used against attributing too much significance to percentile ranks as this test was normed on small, specific groups of children.

In Level III, in the Vocabulary test the stem is a definition, a descriptive phrase, or a word suggesting the response; the task for the children is to find the correct word among different choices. In Speed of Comprehension the task is to find a word omitted from the sentence in the stem, the items are arranged in cycles of similar difficulty, and the test is highly speeded. The Level of Comprehension is tested by questions based on paragraphs of generally increasing difficulty.

Careful note must be made that the post-test percentiles discussed in the following tables are related to the mean raw scores for the post-test. The post-test, as stated earlier, was given to the children during the spring semester of the fifth grade. With children scoring well below the national norms, we confront two problems: 1) did they improve at all during the program year, and 2) if they made progress, did they improve as much as the national sample. In order to answer the first question, the mean score of the post-test was converted into percentile ranking at the same level as the pre-test (i.e., fifth grade). To answer the second question, the mean raw score must be compared with that of the sixth grade national sample.

1) Vocabulary

Table 72 contains the information about the means and standard deviations for the pre and post-test, the correspondent percentile ranks for the means and the t-test value for the mean gain on the Vocabulary subtest of the Reading test in English.

Table 72: Vocabulary: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level III for the Sixth Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles			t-test[2] Values
				Pre-test 5th Grade	Post-test 5th Grade	Post-test 6th Grade	
N	40	22					
Mean	14.00	16.09	2.09	47	55	44	1.23
Standard Deviation	6.30	6.58					

[1]Test scores are raw scores.

[2]t-test Value = 1.23 which was not significant at the .05 level.

From this table we have the following information:

- Children in school C.E.S. 235 finished fifth grade with a mean score in English vocabulary that was slightly below the 50th percentile for native English speaking fifth grade children, as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of 42 percent of fifth grade students from a national sample. This result was expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the English Vocabulary score.

- There were a variety of scores within the group as indicated by the size of the Standa deviation for the pre-test scores. A range of 25 points would be needed in order to include 96 percent of the scores in the sample; 45 is the possible range of scores.

- During eleven months of school, children improved in their English vocabulary scores. The mean score increased by 28 percentile points.

- Children in school C.E.S. 235 did not improve as much as the national sample. Children from our sample finished fifth grade 3 percentile points below the level of a national fifth grade sample; but at the end of the sixth grade they were 6 percentile points below the national level, thus showing an improvement smaller than what is expected from a sixth grade national sample.

- The mean gain of 8 percentile points was not significant beyond the .05 level as indicated by the t-test value. The improvement in English Vocabulary was smaller than what could be expected by chance alone; it was not a real improvement in the trait measured by the test.

- There was a small increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores in English Vocabulary at the end of fifth grade and they had a similar degree of heterogeneity at the end of sixth grade.

2) Speed of Comprehension

Table 73 contains the pertinent information related to the Speed of Comprehension subtest. From this table we conclude that:

Table 73: Speed of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level III for the Sixth Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles			t-test[2] Values
				Pre-test 5th Grade	Post-test 5th Grade	Post-test 6th Grade	
N	42	22					
Mean	10.80	11.36	.56	60	63	46	.35
Standard Deviation	6.40	5.20					

[1] Test scores are raw scores.

[2] t-test Value = .35 which was not significant at the .05 level.

- Children in school C.E.S. 235 finished fifth grade with a mean score in Speed of Comprehension that was above the 50th percentile for native English speaking fifth grade children, as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of 60 percent of fifth grade students from a national sample. This result was not expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the Speed of Comprehension score.

- There were a variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 26 points would be needed in order to include 96 percent of the scores in the sample; 30 is the possible range of scores.

- During eleven months of school, children improved in their Speed of Comprehension scores. The mean score increased by 3 percentile points.

- Children in school C.E.S. 235 did not improve as much as the national sample. Children from our sample finished fifth grade above the level of a national fifth grade sample; but at the end of the sixth grade they were 4 percentile points below the national sixth grade level thus showing an improvement substantially smaller than what is expected from a national sample.

- The mean gain of 3 percentile points was not significant at the .05 level as indicated by the t-test value. The improvement in Speed of Comprehension was smaller than what could be expected by chance alone; it was not a real improvement in the trait measured by the test.

- There was a decrease in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores at the end of fifth grade and they had a smaller degree of heterogeneity at the end of sixth grade.

3) Level of Comprehension

Table 74 has the pertinent information about the Level of Comprehension subtest. The following conclusions are drawn from this table:

Table 74: Level Of Comprehension: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level III for the Sixth Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles			t-test[2] Values
				Pre-test 5th Grade	Post-test 5th Grade	Post-test 6th Grade	
N	42	22					
Mean	11.10	17.13	6.03	22	45	36	4.08
Standard Deviation	4.60	7.21					

[1]Test scores are raw scores.

[2]t-test Value = 4.08 which was significant beyond the .05 level.

- Children in school C.E.S. 235 finished fifth grade with a mean score in Level of Comprehension that was below the 50th percentile for native English speaking fifth grade children, as reported by the pre-test scores. The mean score for our sample is as high or higher than the scores of only 22 percent of fifth grade students from a national sample. This result was expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the Level of Comprehension score.

- There were a variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 18 points would be needed in order to include 96 percent of the scores in the sample; 50 is the possible range of scores.

- During eleven months of school, children improved in their Level of Comprehension scores. The mean score increased by 23 percentile points.

- Children improved more than the national sample. At the end of fifth grade our sample children had a mean score that was at the 22nd percentile of a fifth grade national sample. At the end of the sixth grade their mean score was at the 36th percentile of the sixth grade national sample. They reduced the gap with a national sample by 14 percentile points.

- The mean gain of 23 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in Level of Comprehension was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a substantial increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores at the end of fifth grade and they had a greater degree of heterogeneity at the end of sixth grade.

4) Total Reading

Table 75 has the information in relation to the Total Reading test. The following conclusions are drawn from this table:

Table 75: Total Reading: Means and Standard Deviations for the Pre-test and Post-test and Percentile Ranks on the Inter-American Series Test of Reading, in English Level III for the Sixth Grade.

	Pre-test[1]	Post-test	Mean Gain	Percentiles			t-test[2] Values
				Pre-test 5th Grade	Post-test 5th Grade	Post-test 6th Grade	
N	40	22					
Mean	36.00	44.59	8.59	39	53	42	2.57
Standard Deviation	15.40	17.02					

[1]Test scores are raw scores.

[2]t-test Value= 2.57 which was significant beyond the .05 level.

- Children in school C.E.S. 235 finished fifth grade with a mean score in Total Reading that was below the 50th percentile for native English speaking fifth grade children, as reported by the pre-test scores. The mean score for our sample is as high or higher than only 39 percent of fifth grade students from a national sample. This result was expected; cultural and language differences between native English speaking children and children in school C.E.S. 235 were expected to produce differences in the Total Reading scores.

- There were a variety of scores within the group as indicated by the size of the standard deviation for the pre-test scores. A range of 62 points would be needed in order to include 96 percent of the scores in the sample; 25 is the possible range of scores.

- During eleven months of school, children improved in their Total Reading scores. The mean score increased by 14 percentile points.

- Children in school C.E.S. 235 improved more than the national sample. Children from our sample finished fifth grade at the 39th percentile level of a national fifth grade sample; but at the end of the sixth grade they scored at the 42nd percentile of the national sixth grade level thus showing an improvement greater than what is expected from a national sample.

- The mean gain of 14 percentile points was significant beyond the .05 level as indicated by the t-test value. The improvement in Total Reading was greater than what could be expected by chance alone; it was a real improvement in the trait measured by the test.

- There was a substantial increase in group variability between pre and post-testing as indicated by the size of the standard deviations. Children were a heterogeneous group according to their scores at the end of fifth grade and they had a greater degree of heterogeneity at the end of sixth grade.

5) Comparison of Subtests

Table 76 includes the information related to each subtest and the total test. From this table we draw the following conclusions:

Table 76: Means and Standard Deviations for the Pre-test, Post-test, Percentile Ranks of each Subtest of the Inter-American Series Test of Reading, in English, Level III, for the Sixth Grade.

Test	Pre-test[1]			Post-test			Mean Gain	Percentiles[2]			t-test[3] Values
	N	Mean	SD	N	Mean	SD		Pre-test 5th	Post-test 5th	Post-test 6th	
Vocabulary	40	14.00	6.30	22	16.09	6.58	2.09	47	55	44	1.23
Speed of Comprehension	42	10.80	6.40	22	11.36	5.20	.56	60	63	46	.35
Level of Comprehension	42	11.10	4.60	22	17.13	7.21	6.03	22	45	36	4.08*
Total Reading	40	36.00	15.40	22	44.59	17.02	8.59	39	53	42	2.57*

[1]Test scores are raw scores.

[2]Percentiles are for Fifth and Sixth Grade.

[3]*Significant beyond the .05 level.

- Children in school CES 235 finished fifth grade, as reported by the pre-test data, with a mean score in Speed of Comprehension, higher than the 50th percentile of the native English speaking fifth grade children. They scored lower than the national sample in Level of Comprehension , Total Reading and Vocabulary.

- There were a variety of scores in every subtest as indicated by the size of the standard deviations. The group was a heterogeneous group and it had similar degrees of heterogeneity on each subtest.

- During eleven months of school, children improved in every subtest. The greatest improvement was in Level of Comprehension

and Total Reading where the difference between pre and post-test mean scores was 23 and 14 percentile points.

- The mean gains were significant beyond the .05 level only in Level of Comprehension and Total Reading as indicated by the t-test values. These mean gains in Vocabulary and Speed of Comprehension were not beyond what could be expected by chance alone and, therefore, we can not say that the gains were real gains in the abilities measured by the subtests.

- Children, during the school year, reduced the gap between their mean score and the 50th percentile for a national sample in Level of Comprehension and Total Reading. At the end of the fifth grade our sample children were 28 and 11 percentile points below the 50th percentile for a national fifth grade; a year later the gap between their mean score and the national sixth grade sample was only 14 percentile points in Level of Comprehension and 8 in Total Reading.

- Children increased the gap between their mean score and the 50th percentile for a national sample in Vocabulary and Speed of Comprehension. Their mean score in Vocabulary was 3 points below the 50th percentile of a national fifth grade and 8 points at the end of the sixth grade, thus increasing the gap by 3 percentile points. The mean score in Speed of Comprehension was 10 points above the 50th percentile of the national fifth grade but at the end of the sixth grade they were 4 points below the 50th percentile of a national sixth grade, thus losing 14 percentile points during the year.

6) Correlation Between Subtests

Table 77 reports the correlations between the three subtests for school C.E.S. 235 and for the national sample.

Table 77: Product-Moment Correlation Coefficients Between Vocabulary, Speed and Level of Comprehension on the Inter-American Series Test of Reading in English, Level III, for the Sixth Grade.

	C.E.S. 235	National Sample
Vocabulary vs. Speed of Comprehension	.52	.80
Vocabulary vs. Level of Comprehension	.19	.86
Speed of Comprehension vs. Level of Comprehension	.65	.83

From this table we can draw the following conclusions:

- The correlation between Vocabulary and Speed of Comprehension and between Speed and Level of Comprehension was moderate. Changes on one of the subtests were followed by moderate changes on the other subtest. About 27 percent of the changes in Speed of Comprehension were related to changes in Vocabulary; about 42 percent of the changes in Level of Comprehension were related to Speed of Comprehension.

- The correlation between Vocabulary and Level of Comprehension was very low. Only about 4 percent of the changes on one of the subtests was related to changes on the other.

- It is an important finding that Level of Comprehension is almost unrelated to Vocabulary. To what, then, is it related? The Level of Comprehension is tested by questions based on paragraphs. Two possible explanations came to our minds: 1) the paragraphs are irrelevant to the students tested and, therefore, their scores do not depend on knowledge of words but on something like attention, retention in memory, interest, and so forth; all factors related to the meaningfulness of the paragraphs; or 2) it could be that the paragraphs were very difficult for these students, and the scores, then, were the result of guessing on most of the items because the children did not understand them. If the scores are the result of chance, they can not be related to anythind and it could explain why they were not related to Vocabulary.

The mean score is very close to the the score which can be obtained by guessing alone; this would suggest that the paragraphs were very difficult. But, on the other hand, Level of Comprehension is moderately related to Speed of Comprehension, suggesting that the scores were not obtained by guessing alone; this would support the idea that the content of the paragraphs was more irrelevant to the students than difficult.

- The correlation coefficients were lower in school C.E.S. 235 than in the national sample. This means that improvement in one of the English language skills is more related to improvements in other language skills in a national sample than in school C.E.S. 235. The different language skills were more independent in our sample than in a native English speaking sample.

c. Tests of Basic Language Competence

Tests of language competence in Spanish and English were given to all children who were judged by the teachers to be minimally competent in Spanish or English. Only subtests 1 and 2 were given. A different form was used in Spanish and in English.

Since these tests have not been normed on a national sample, percentile ranks are not included here. The only possible comparison is to a perfect total score, which are reported at the bottom of each table.

The test of Language Competence was designed to measure a child's ability to understand Spanish and English. Children who speak one or both languages as a native are expected to make near perfect scores on the subtests.

The subtest 1, Oral Vocabulary, presents words to the child from which the child must identify the corresponding picture. The subtest 2, measures phonemic discrimination at word and phrase level; the test requires the child to discriminate whether two words sound the same or different.

1) Oral Vocabulary

Table 78 has the pertinent information about the Oral Vocabulary in Spanish and in English. From the data in this table we may conclude that:

Table 78: Oral Vocabulary Means and Standard Deviations on Spanish and English Versions of the Language Competence Test for the Sixth Grade.

Language	Pre-test[1]			Post-test			Mean Gain	t-test Values[2]
	N	Mean	SD	N	Mean	SD		
Spanish				25	23.12	2.90		
English	23	21.90	3.30	22	23.63	1.29	1.73	2.30*

[1]Test scores are raw scores.

[2]*Significant beyond the .05 level.

NOTE: Perfect scores = 24.

- Children from school C.E.S. 235 scored low in English in the pre-test. Children finished fifth grade with a knowledge of English words that was less than perfect.
- Students form a heterogeneous group according to their scores in English in the pre-test. The highest possible range is 24 but a range of more than 13 is needed in order to include 96 percent of the scores in English.
- Children changed significantly in English Oral Vocabulary during eleven months of school. The positive mean change of 1.78 between pre and post-test was significant and it could not be the result of chance alone.
- Students scored near a perfect score in Spanish at the end of the sixth grade. Their mastery of simple Spanish words was almost perfect, as a group.
- Children became a more homogeneous group in their English scores during the school year as indicated by a substantial decrease in the size of the standard deviation between the pre and post-testing.

2) Phonemic Discrimination

Table 79 reports the pertinent information about the subtest Phonemic Discrimination in Spanish and in English. From this table we have the following information:

Table 79: Phonemic Comprehension: Means and Standard Deviations on Spanish and English Versions of the Language Competence Test for the Sixth Grade.

Language	Pre-test[1]			Post-test			Mean Gain	t-test Values[2]
	N	Mean	SD	N	Mean	SD		
Spanish	28	34.90	4.00	22	19.81	2.15	-15.09	15.95*
English	22	33.90	6.80	22	36.86	2.60	2.96	1.91*

[1]Test scores are raw scores.

[2]*Significant beyond the .05 level.

NOTE: Perfect score = 40.

- Children from school C.E.S. 235 scored low in Spanish and in English in the pre-test. Their mean score at the end of second grade was less than perfect; their mastery of Spanish and English sounds was less than expected from children that speak either language as a native language. Students from our sample had a language disadvantage in relation to both Spanish and English speaking students.

- Children had similar discrimination of English and Spanish sounds as shown by the small difference in mean score for the pre-test.

- Students were a heterogeneous group according to their scores in Spanish and in English as indicated by the size of the standard deviation for the pre-test. There were a wide range of scores in both languages.

- During eleven months of school children decreased their mean score in Spanish as shown, by the negative mean gain. The negative change, was significant beyond the .05 level only in English. The change was a real negative change indicating a deterioration of their ability to discriminate sounds.

- There was a substantial decrease in group heterogeneity in Spanish and in English scores; students, as a group, sharply reduced their differences.

- Children increased their mean score in English during the school year. The mean gain was significant beyond the .05 level and it represented an improvement in ability to discriminate English sounds which could not be the effect of chance alone. The students, though, as a group did not have a perfect discrimination of English sounds at the end of the sixth grade; the mean score was not a perfect or near a perfect score.

B. CLASSROOM STRUCTURE, CONTENT AND LANGUAGE

1. Evaluation Methodology

Forty-five classes were observed for thirty minutes each in order to obtain information about the instructional materials, content and language of instruction, bicultural aspects of the lessons, classroom organization, use of language by students and adults, and instructional processes by bilingual observers from Mobicentrics. Each of the major areas of information will be discussed separately after an analysis of the general characteristics of the teachers and classes observed.

2. General Characteristics of the Classroom

Table 80 presents the general characteristics of the teachers and classes observed.

Table 80: General Characteristics of the Classroom

Grade	No. Of Classes	Mean Size	Children Speaking Predominantly Spanish*		Children Speaking Predominantly English*		Teachers[1] Speaking Predominantly Spanish		Teachers Speaking Predominantly English		Paraprofessionals[2] Speaking Predominantly Spanish		Paraprofessionals Speaking Predominantly English	
			N	%	N	%	N	%	N	%	N	%	N	%
Kindergarten, 1st, and 2nd grade	18	13.9	105	42	146	58	6	33	12	67	5	50	5	50
3rd and 4th grade	15	18.1	90	31	196	69	3	23	10	77	3	75	1	25
5th and 6th grade	12	23.3	74	26	210	74	5	38	8	62	5	56	4	44
TOTAL	45	17.8	269	33	552	67	14	33	30	68	13	57	10	43

[1]Teacher totals do not always add to the total number of observations because in two cases, no teachers were present and in one case two teachers were present.

[2]Every class did not have a paraprofessional at the time of observation.

Of the 45 classes observed, 18 were kindergarten, first, or second grades, 15 were third and fourth grades, and 12 were fifth or sixth grades. The mean class size for kindergarten, 1st and 2nd grades was 13.9 pupils, for 3rd and 4th grades it was 18.1, and for 5th and 6th grades it was 23.3. The size of class increased as the children got older. The overall mean size of classes observed was 17.8.

The dominant language of sixty-seven percent of all children observed was English; the dominant language for the other thirty-three percent was Spanish. As the children became older there was a trend toward an increased precentage of children whose dominant language was English. The dominant language of fifty-eight percent of the children in kindergarten, first and second grade was English; English was the dominant language of seventy-four percent of fifth and sixth graders, an increase of sixteen percent. In sixty-eight percent of 45 observations, the dominant language of the teachers observed was English. This percentage remained relatively constant for the teachers in all classes. However, in fifty-seven percent of the observations the dominant language of the paraprofessionals was Spanish. There were too few paraprofessionals observed to break down the dominant language for them by grade. It appears that the district made an extensive effort to hire paraprofessionals whose dominant language was Spanish for this program, even though the dominant language of the majority of the students was English. This decision makes sense if one assumes that Spanish dominant language adults can best assist Spanish dominant language children especially since, the majority of teachers have a dominant language of English. It is assumed all paraprofessionals and teachers are bilingual in this bilingual program and this data reflects only the dominant language of children, teachers and aides.

3. Instructional Materials in the Classroom

The observer evaluated materials in the classroom visited in ten areas on a 1 = none in evidence to 4 = very large numbers in evidence scale. The observer found the following:

- 43 (96%) of the 45 classes had at least some evidence of blackboards.
- 38 (84%) of the 45 classes had at least some evidence of libraries.
- 27 (60%) of the 45 classes had at least some evidence of display of student's work.
- 35 (78%) of the 45 classes had at least some evidence of school made classroom exhibits.

- 27 (60%) of the 45 classes had at least some evidence of a variety of other classroom exhibits.
- 19 (42%) of the 45 classes had at least some evidence of materials dealing with Hispanic Cultures.
- 19 (42%) of the 45 classes had at least some evidence of materials dealing with Anglo American Culture.
- 14 (31%) of the 45 classes had at least some evidence of audio visual equipment.
- 38 (84%) of the 45 classes had at least some evidence of manipulative equipment.
- 39 (87%) of the 45 classes had at least some evidence of textbooks.

In general the classrooms had ample materials. However, for a bilingual school there were a surprising number of classrooms (58%) that showed no evidence of either Hispanic or Anglo American culture materials. Additionally, although more than fifty percent of the classes showed no evidence of audio visual equipment, this equipment might have been stored in a central supply room or in closets. The above data relates only to materials on display.

4. Content and Language of Instruction

In this section the observer was trying to determine the amount of time spent teaching in English and Spanish and the amount of time devoted to each subject. Of the 900 minutes of teaching observed (20 minutes for 45 classes), 380 minutes or 42% of the time, the teacher taught in English; 58% of the time was taught in various combinations of English and Spanish. In other words, in bilingual education classes about three-fifths of the time was spent teaching bilingually. One hundred sixty-five or 18% of the 900 minutes observed was spent in teaching Spanish language arts; 300 minutes or 33% of the time observed was spent in teaching English language arts; and 435 minutes or 48% of the time observed was spent teaching math, social studies, science or other subjects. It seems reasonable that about 50% of the time should be spent on teaching other subjects than English or Spanish. However, if the goal of the school is to eventually make all children bilingual and about 2/3 of the children have a predominant language of English, then it seems to be more reasonable to be spending much more of the language time on Spanish language arts instead of English as now is the case. It seems it should take at least the same amount of time to teach English speaking children how to speak Spanish as it does to teach Spanish speaking children how to speak English. The modifying factor in the time allotment in the teaching of Spanish and English

language arts is the fact that fifty-eight percent of the total teaching time is spent in some combination of English and Spanish languages. Therefore, the students are likely getting some Spanish in the math and other subject matters along with the 18% they get in formal Spanish language instruction.

5. Organization of the Classroom

The purpose of this part of the instrument was to determine how the class was organized when it was taught.

Table 81 shows the breakdown of the number of minutes and percentage of time spent in each class structure.

Table 81: Percentage Time in Minutes Spent in Intact, Small Group and Individual Instruction by Classes With and Without Paraprofessionals.

	Intact		Small Group		Individual Instruction		Total	
	Min	%	Min	%	Min	%	Min	%
Classes With Paraprofessionals	170	39	220	51	45	10	435	100
Classes Without Paraprofessionals	275	38	385	53	60	9	720	100
TOTAL	445	39	605	52	105	9	1155	100[1]

[1]Because some classes had more than one structure occurring at the same time, the total number of minutes is over the 900 minutes of observation time.

Of the classes observed, 445 minutes or 39% of the time occurred in an intact class; 605 minutes or 52% of the time occurred in small groups; and 105 minutes or 9% of the time occurred in individual instruction. Even when these totals were broken down into those 16 classes who had paraprofessionals and compared their classroom organization with the 29 classes who did not have paraprofessionals, it was found that the percentages stayed roughly the same. That is, contrary to what one might expect; having paraprofessionals did not change how students were organized in the classroom.

6. Use of Language By Students and Adults in the Classroom

In eighteen (41%) out of forty-five classes noted, the majority of the students spoke entirely in English. Also the teachers of eighteen (44%) out of forty-one classes

spoke entirely in English during the observation. In 14% of the observations the majority of students spoke entirely in Spanish and in 22% of the observations the teachers spoke entirely in Spanish. In the other 46% of the observations on students they spoke mostly in English 23% of the time, half in Spanish, half in English 9% of the time and mainly in Spanish 14% of the time. The teachers spoke mainly in English 12% of the time half in Spanish, half English 10% of the time and mainly in Spanish 12% of the time.

The dominant language used by teachers was English since for 64% of the observations they spoke mainly or entirely in English. This was true for students as well as they spoke mainly or entirely in English 56% of the time. On the other hand, only 3 (15%) of the twenty aides observed spoke entirely in English and only 30% (6) of them spoke mainly in English. They spoke entirely or mainly in Spanish 45% of the time. This gives further support to the conclusion that the paraprofessionals were likely to help the Spanish speaking pupil.

7. Instructional Processes

Nineteen areas of concerned with how pupils and teachers interacted were observed and rated on a scale of 1 = not at all, 2 = a small amount of time, 3 = about half the time, 4 = most of the time, and 5 = all the time. The results for all the nineteen areas were:

- Of the 45 classes observed for the teacher talking while children listened the teacher talked half the time or more in all, but one (98%) of them. The modal level for the teacher talking in the classes was "most of the time".

- Of the 44 classes observed for this issue, the teacher provided opportunity for children to talk about what they were doing in class half the time or more in all but one of them (98%). The modal level of teachers providing children with opportunity for talking about what they were doing in class was "all of the time."

- The pupils were provided an opportunity to interact vervally in their second language half the time or more in 18 (40%) of the classes. The modal level of pupils interacting verbally in their second language was between "a small amount of time" and "about half the time".

- Of the 43 classes observed, the leader used physical activity half the time or more in 27 (60%) of them. The modal level of physical activity was between about half of the time and "most of the time".

- There was positive physical contact between teacher and pupils half the time or more in 40 (93%) of the classes. The modal level of positive physical contact was between "most of the time" and all of the time.

- The teacher employed motivational techniques in introducing materials to students half the time or more in 98% of the classes. The modal level of motivational techniques was between "most of the time" and "all of the time".

- The teacher used feedback systems with pupils in order to let them know how they were performing various learning tasks, half the time or more in 98% of the classes. The modal level of use of feedback systems was between "most of the time" and "all of the time".

- Teachers used a range of specific techniques for dealing with disruptive pupils in the classroom half or more of the time in 88% of the classes. The modal use of a range of discipline techniques was "most of the time".

- There was a variation in the presentation of subject matter according to varying levels of pupil development, accomplishment, motivation, interest or modes of learning half or more of the time in 41 (95%) of the classes. The modal use of variation in the presentation of subject matter was "most of the time".

- The teacher attempted to foster an atomsphere or attitude of "self dependence" rather than one of "other dependence" half or more of the time in all of the classes. The modal attempt to foster "self dependence" was "most of the time".

- Pupils initiated their own instructional activity half or more of the time in 68% of the classes. The modal pupil initiation was between "about half of the time" and "most of the time".

- The teacher encouraged pupils to use adults other than herself as resource persons half or more of the time in (65%) of the classes when another adult was present. The modal encouragement to use other adults was "about half the time".

- The auxiliary personnel took an active and substantial role in the processes of classroom instruction and learning half or more of the time in (67%) of the classes. The modal encouragement to use other adults was between "about half the time" and "most of the time"

- The teacher indicated an attitude of "the school cares" to the pupils half or more of the time in all of the classes. The modal attitude of "the school cares" was "all of the time".

- The teacher generally related to pupils in ways which demonstrated acceptance and respect for the individual child's native language and culture, national origin, race, family and social or economic background half or more or the time in all (100%) of the classes. The modal demonstration of acceptance and respect for the individual child's native language etc., was "all the time".

- When the teacher was observed using a second language, her "use of second language approximated a native speaker's norm with respect to pronunciation, syntax, vocabulary and semantic" half or more of the time in all of the classes. The modal use of a second language by teachers which approximated a native speaker was "all the time".

- The auxiliary personnel's use of a second language approximated a native speaker norm with the respect to pronunciation, syntax, vocabulary and semantics half or more of the time in all classes. The modal use of a second language by auxiliary personnel which approximated a native speaker was between "most of the time" and "all of the time".

- The teacher handled efficiently and with dispatch the non-instructional processes of the classroom half or more of the time in all but one (98%) of the classes. The modal time of handling efficiently non-instructional processes was "most of the time".

- The students responded to the teacher in the same language as they were addressed half or more of the time in all (100%) of the classes. The modal of responding in the same language was between "most of the time" and "all the time".

To summarize the observations, several things appear from the data. First, the teachers and students appear to be well matched in terms of dominant language. In 68% of the observations, teachers were English dominant and 67% of the children observed were English dominant. Secondly, the classrooms in general had a large quantity of instructional materials present in the classroom. There were displays of student's work in 60% of the classrooms; textbooks, libraries, blackboards, classroom exhibits, and manipulative equipment were present in at least 78% of the classrooms. The only materials which seemed to be missing from the classrooms were materials dealing with Hispanic Culture (present in 47% of the classes) and materials dealing with Anglo American Culture. It should be noted, however, that these materials are present in the school corridors.

The classes, true to the program goals, were organized primarily in small groups. Over half the time observed was spent in this structure.

In contrast to last year when the observers noted only a small percentage of time spent teaching in Spanish, this year 58% of the time was taught in a combination of Spanish and English. Although both teachers and students spoke mainly or entirely in English a large percent of the time, they also spoke mainly or entirely in Spanish a fair amount of the time. The school appears to clearly be moving toward a truly bilingual environment.

In addition, the classroom appeared to be operating with a great deal of feedback and communication with students, one goal of the program. Students were also getting the opportunity to express themselves in both Spanish and English.

C. STAFF DEVELOPMENT

1. Evaluation Methodology

Interviews with the program coordinator and school records formed the basis for the following information.

2. Teacher and Paraprofessional Demographics

All 15 teachers were bilingual and spoke both English and Spanish fluently, however, most of them had a dominant language which they preferred to speak. For 4 of the 15 teachers, this was English; for 4, Spanish and 7 were truly bilingual and equally comfortable in either language. Of the 6 paraprofessionals, 1 was English dominant, 4 were Spanish dominant and 1 was completely bilingual.

Four teachers had lived in a Spanish speaking community while eleven teachers had lived in both English and Spanish speaking communities for at least five years. None had lived in English dominant communities only.

Seven teachers received their formal education in English while eight received their formal education in both Spanish and English.

3. Teacher Training

All teachers were trained both prior to school opening in January, 1972 and during the 1972 school year.[1] In addition, new staff members received training before the school opened in September, 1972. They received three days of training in teaching reading in both languages, 3 days in teaching math and two days in teaching a second language. The training sessions were held 8 hours a day, for a total of eight days. The average attendance at the training sessions was 97%.

Fourteen teachers were enrolled in bilingual programs at universities; seven teachers took 78 credits at N.Y.U., three teachers took 21 credits at Fordham, one teacher each took courses at Lehman and City College and two teachers were enrolled at Hunter. Two paraprofessionals were also taking college courses. These courses involved the study of "American English", Spanish language, dramatics, bilingual teaching, curriculum and child development as well as a variety of other

[1] See Evaluation of Title VII, Bilingual Educttion Program, Community School District 9, Bronx, N.Y., 1971-1972, by Mobicentrics, Inc.

subjects.

In addition to pre-service training, sessions were conducted at least once a month for in-service training. Group training was conducted in two hour monthly sessions, while individual and small group sessions generally lasted from 45 minutes to two hours.

4. Staff Meetings

Monthly staff meetings for all teachers were held in addition to in-service training meetings. The average attendance at each meeting was between 90 and 100%. Each meeting lasted approximately 2 hours and the general break down of meetings was that approximately 50% of the time was spent on planning for the following month and 25% each was spent on discussions of past plans and open discussions. In addition meetings were frequently held with small groups of teachers. Special meetings were held with teachers in the team teaching group, the Spanish teachers, grade teachers, and each grade teacher's paraprofessionals. These meetings included such areas as reorganization plans, special teaching units around certain holidays such as Puerto Rican Discovery Week, Martin Luther King Day, Christmas, second language instruction, graduation and curriculum development.

D. COMMUNITY ACTIVITIES

1. Evaluation Methodology

Interviews with the Program Director and analysis of school records and newsletter constitutes the information in this section.

2. Bilingual Newsletter

The bilingual newsletter was published approximately once a month from January, 1973. Approximately 400 copies of the newsletter were published with 350 going to parents of children in the school, 25 to community groups and 25 to other persons who picked them up in the office. The copies to parents were sent through the children, those to community groups by mail and the rest were picked up by those who wanted them. The content of the newsletter dealt with general information about the school (50%) educational articles (25%) and other topic areas such as community news and events.

A sample newsletter emphasized parent-teacher cooperation, and covered such areas as grade level conferences Open School Week, candy sales, voting for the Community School Board, and the annexation of Loew's 167th Street Theater. The letters were very clear and warmly written.

3. Parent's Association

There were approximately 75 parents actively involved in the P.A. at CES 235. Sixty of these parents attended the monthly meetings which last two to three hours. The meetings helped to inform the parents about school activities, get parents opinions about school problems and policies and organized the parents to take action with the school board.

4. Other Activities

The community was also brought into the school in a variety of other ways. Through special activities centered around Puerto Rican Discovery Week and Martin Luther King Day the community had a chance to become involved with the school.

E. MATERIALS

1. Evaluation Methodology

An interview with the Program Director and analysis of school materials and purchase orders yielded the data for the following section.

2. Professional Staff Library

Information received on the materials used, indicated the existence of a professional Bi-lingual Library for staff with a large number of books in Spanish (274), the major content area being Hispanic Culture (206). A very small percentage of books included the Sciences, American Culture, and Language (66 books). In addition, 132 English books were in the library. Copies of purchase orders to the Puerto Rican Heritage Publications showed the purchase of a large number of all Spanish materials. It would seem, on the basis of materials received, that there should be a larger number of English, and Spanish-English materials to achieve a goal of bi-lingualism and bi-culturalism.

3. Exchange of Materials With Sister School

No mechanism for exchanging materials with the Sister School in Puerto Rico has yet been affected.

V. RECOMMENDATIONS

- More emphasis needs to be placed on English Vocabulary development in every grade except sixth.

- More emphasis also needs to be placed on reading skills in all grades. Children are deficient in many reading skills.

- In addition, Speed of Comprehension and Vocabulary need to be improved upon in both languages.

- The objective of exchanging materials with the sister school in Puerto Rico was never met. A good way of helping children learn Spanish language skills could be to initiate a letter exchange between the two schools.

- A greater effort should be made in making the classroom environment reflect the bilingual culture. Pictures, classwork, maps and other visual aids, properly labeled, could act as cues to the students to increase their vocabulary in both languages. Only 42% of the classes observed attempted this technique.

- More time should be spent on teaching the Spanish language. Twice as much time is spent on teaching English than is spent on teaching Spanish even though the dominant language of 2/3 of the pupils is English. In addition, scores on the tests in Spanish Vocabulary were considerably lower than English.

- Greater use of paraprofessionals to increase the flexibility of the classroom organization should be attempted. Currently having paraprofessionals in the classroom does not significantly increase small group or individual instruction.

- In general, there is a substantial improvement over last year in almost every area. The evaluators recommend that the program definitely be recycled next year, as the program appears to be achieving beneficial effects on the students involved. The school appears to be moving toward a truly bilingual open classroom situation.

mobicentrics inc.
384 east 149th st.
bronx, n.y. 10455